



Strict Limits on Potential Secondary Atmospheres on the Temperate Rocky Exo-Earth TRAPPIST-1 d

Caroline Piaulet-Ghorayeb^{1,2,24} , Björn Benneke^{1,3} , Martin Turbet^{4,5} , Keavin Moore^{6,7} , Pierre-Alexis Roy¹ , Olivia Lim¹ , René Doyon¹ , Thomas J. Fauchez^{8,9,10} , Loïc Albert¹ , Michael Radica^{1,2,25} , Louis-Philippe Coulombe¹ , David Lafrenière¹ , Nicolas B. Cowan^{6,7} , Danika Belzile^{11,12} , Kamrul Musfirat¹³ , Mehramat Kaur¹⁴ , Alexandrine L'Heureux¹ , Doug Johnstone^{15,16} , Ryan J. MacDonald^{17,26} , Romain Allart¹ , Lisa Dang¹ , Lisa Kaltenegger^{18,19,20} , Stefan Pelletier^{1,21} , Jason F. Rowe²² , Jake Taylor^{1,23} , and Jake D. Turner^{18,20}

¹ Trottier Institute for Research on Exoplanets and Department of Physics, Université de Montréal, Montréal, QC, Canada; carolinepiaulet@uchicago.edu

² Department of Astronomy & Astrophysics, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA

³ Department of Earth, Planetary, and Space Sciences, University of California, Los Angeles, CA 90095, USA

⁴ Laboratoire de Météorologie Dynamique/IPSL, CNRS, Sorbonne Université, Ecole Normale Supérieure, PSL Research University, Ecole Polytechnique, 75005 Paris, France

⁵ Laboratoire d'astrophysique de Bordeaux, Université de Bordeaux, CNRS, B18N, allée Geoffroy Saint-Hilaire, 33615 Pessac, France

⁶ Department of Physics, McGill University, Montréal, QC, Canada

⁷ Department of Earth and Planetary Sciences, McGill University, Montréal, QC, Canada

⁸ NASA Goddard Space Flight Center, 8800 Greenbelt Road, Greenbelt, MD 20771, USA

⁹ Integrated Space Science and Technology Institute, Department of Physics, American University, Washington DC, USA

¹⁰ NASA GSFC Sellers Exoplanet Environments Collaboration, USA

¹¹ Department of Physics, Université de Montréal, Montréal, QC, Canada

¹² Dawson College, Montréal, QC, Canada

¹³ Department of Physiology, McGill University, Montréal, QC, Canada

¹⁴ École secondaire Cavalier-De LaSalle, Montréal, QC, Canada

¹⁵ NRC Herzberg Astronomy and Astrophysics, 5071 West Saanich Road, Victoria, BC V9E 2E7, Canada

¹⁶ Department of Physics & Astronomy, University of Victoria, Victoria, BC V8P 5C2, Canada

¹⁷ Department of Astronomy, University of Michigan, 1085 South University Avenue, Ann Arbor, MI 48109, USA

¹⁸ Carl Sagan Institute, Cornell University, Ithaca, NY, USA

¹⁹ Cornell Center for Astrophysics and Planetary Science, Cornell University, Ithaca, NY 14853, USA

²⁰ Astronomy Department, Cornell University, Ithaca, NY 14853, USA

²¹ Observatoire astronomique de l'Université de Genève, 51 Chemin Pegasi, 1290 Versoix, Switzerland

²² Department of Physics and Astronomy, Bishop's University, 2600 Rue College, Sherbrooke, QC J1M 1Z7, Canada

²³ Department of Physics, University of Oxford, Parks Road, Oxford OX1 3PU, UK

Received 2025 February 27; revised 2025 July 7; accepted 2025 July 17; published 2025 August 13

Abstract

The nearby TRAPPIST-1 system, with its seven small rocky planets orbiting a late-type M8 star, offers an unprecedented opportunity to search for secondary atmospheres on temperate terrestrial worlds. In particular, the $0.8 R_{\oplus}$ TRAPPIST-1 d lies at the edge of the habitable zone ($T_{\text{eq,A}=0.3} = 262 \text{ K}$). Here we present the first $0.6\text{--}5.2 \mu\text{m}$ NIRSpec/PRISM transmission spectrum of TRAPPIST-1 d from two transits with JWST. We find that stellar contamination from unocculted bright heterogeneities introduces 500–1000 ppm visit-dependent slopes, consistent with constraints from the out-of-transit stellar spectrum. Once corrected, the transmission spectrum is flat within $\pm 100\text{--}150 \text{ ppm}$, showing no evidence for a haze-like slope or molecular absorption despite NIRSpec/PRISM's sensitivity to CH_4 , H_2O , CO , SO_2 , and CO_2 . Our observations exclude clear, hydrogen-dominated atmospheres with high confidence ($>3\sigma$). We leverage our constraints on even trace amounts of CH_4 , H_2O , and CO_2 to further reject high mean molecular weight compositions analogous to a haze-free Titan, a cloud-free Venus, early Mars, and both Archean Earth and a cloud-free modern Earth scenario ($>95\%$ confidence). If TRAPPIST-1 d retains an atmosphere, it is likely extremely thin or contains high-altitude aerosols, with water cloud formation at the terminator predicted by 3D global climate models. Alternatively, if TRAPPIST-1 d is airless, our evolutionary models indicate that TRAPPIST-1 b, c, and d must have formed with $\lesssim 4$ Earth oceans of water, though this would not preclude atmospheres on the cooler habitable-zone planets TRAPPIST-1 e, f, and g.

Unified Astronomy Thesaurus concepts: Extrasolar rocky planets (511); Exoplanet atmospheres (487); Exoplanets (498); M dwarf stars (982); Stellar activity (1580); Starspots (1572); Stellar faculae (1601); Transmission spectroscopy (2133); Habitable planets (695)

Materials only available in the online version of record: machine-readable table

²⁴ E. Margaret Burbridge Postdoctoral Fellow.

²⁵ NSERC Postdoctoral Fellow.

²⁶ NHFP Sagan Fellow.

1. Introduction

TRAPPIST-1 is a nearby (12 pc) small ($0.12 R_{\odot}$, see Table A1; E. Agol et al. 2021) ultracool M8V star (J. Liebert & J. E. Gizis 2006; M. Gillon et al. 2017; F. Davoudi et al. 2024), host to seven transiting rocky planets receiving moderate irradiation (M. Gillon et al. 2017). The extremely small size of the star boosts the detectability of even high mean

molecular weight secondary atmospheres (C. V. Morley et al. 2017; A. P. Lincowski et al. 2018; J. Lustig-Yaeger et al. 2019) on top of the small TRAPPIST-1 planets (approximately $0.75\text{--}1.1 R_{\oplus}$), and shrinks the orbital distance where water can be in the liquid form (i.e., the habitable zone (HZ)) relative to Sun-like stars. This offers frequent opportunities to probe the atmospheres of potentially habitable planets on orbital periods of only about 4–10 days.

TRAPPIST-1 d is the third planet from the host star. At an orbital distance of a mere 0.02 au, the exo-Earth ($0.8 R_{\oplus}$, see Table A1) receives only 1.12 times as much flux from its host star as Earth receives from the Sun (M. Gillon et al. 2017). Because of its proximity to its host star, however, TRAPPIST-1 d is believed to be tidally locked with a permanent night- and dayside (M. Turbet et al. 2020). TRAPPIST-1 d is predicted by 1D models to lie inside the runaway greenhouse limit (R. K. Kopparapu et al. 2013; M. Gillon et al. 2017; H. Massol et al. 2023). However, 3D global climate models (GCMs) predict that the radiative feedback and reflectivity of dayside clouds may allow for liquid water conditions at its surface given its synchronously rotating state (J. Yang et al. 2013; R. K. Kopparapu et al. 2016), if cloud properties are favorable (E. T. Wolf 2017; M. Turbet et al. 2018; M. J. Way 2025). The presence of surface liquid water, however, presupposes that TRAPPIST-1 d cooled down enough after its formation to allow for water to condense out of the atmosphere, which is unlikely at its present-day orbital distance (M. Turbet et al. 2023). With its short (4 day) orbital period, which enables many repeat observations, and its low mass ($0.39 M_{\oplus}$, compared to 0.69 and $1.04 M_{\oplus}$ for the outer planets e and f, respectively), TRAPPIST-1 d is an extremely favorable small temperate planet for transmission spectroscopy (M. L. Hill et al. 2023). Overall, this planet stands out as a great candidate for an observational test of the HZ concept, as different model assumptions lead to a range of predictions about its composition and surface conditions.

The seven planets’ near-resonant chain produces transit-timing variations (S. Wang et al. 2017; S. L. Grimm et al. 2018; E. Agol et al. 2021) that enable precise mass measurements, which would be challenging to obtain using radial velocities. The latest determination of planet masses and radii in the TRAPPIST-1 system demonstrated that, within measurement uncertainties, all seven planets can be explained by the same iso-composition curve (E. Agol et al. 2021). They, however, have consistently lower bulk densities than the solar system terrestrial planets, indicating either lower amounts of iron in their interior or moderate enrichment in volatile species such as water. The TRAPPIST-1 planets could even be “core-free” planets (L. T. Elkins-Tanton & S. Seager 2008), where all the iron has oxidized and remains within the mantle rather than segregated in a metal core. While the planets’ bulk densities are similar, they show a tentative trend of lower density with increasing stellocentric distance, suggesting a gradient in volatile endowment. In this trend, TRAPPIST-1 d stands out as an outlier, with a lower density than expected for a composition consistent with planets c and e. This leads to varying estimates of the water mass fraction of planet d, from less than 0.01% assuming water is in the vapor form (E. Agol et al. 2021), and up to a few percent assuming the surface conditions allow for liquid water to form (L. Acuña et al. 2021). Such volatiles could have been provided by a formation near the ice line, through pebble accretion (C. T. Unterborn et al. 2018;

G. A. L. Coleman et al. 2019), or if it formed in a “cold finger” of the protoplanetary disk where water ice was locally more abundant (K. E. Cyr et al. 1998).

Recent explorations with JWST of the two innermost TRAPPIST-1 planets, TRAPPIST-1 b (J. Ih et al. 2023; O. Lim et al. 2023; T. P. Greene et al. 2023; E. Ducrot et al. 2024) and TRAPPIST-1 c (S. Zieba et al. 2023; A. P. Lincowski et al. 2023; M. Radica et al. 2025), revealed poor heat redistribution (in emission) and flat spectra (in transmission), to the level of excluding some high mean molecular weight scenarios, overall favoring a bare-rock interpretation. Another key finding from atmosphere searches on the TRAPPIST-1 planets (see, e.g., J. de Wit et al. 2018; Z. Zhang et al. 2018; H. R. Wakeford et al. 2019; L. J. Garcia et al. 2022; O. Lim et al. 2023; M. Radica et al. 2025; A. D. Rathcke et al. 2025) is that potential planetary molecular absorption features in the measured transmission spectra are overwhelmed by 200–300 ppm amplitude signals that are varying from one visit to the next, and are likely caused by the transit light source (TLS) effect. The TLS effect is caused by the presence of patches of the unocculted stellar surface (e.g., spots or faculae) that are cooler or hotter than the photosphere along the planet’s transit path (D. K. Sing et al. 2011; Z. K. Berta et al. 2011; B. V. Rackham et al. 2018). Although this effect can be marginalized over (see, e.g., M. Fournier-Tondreau et al. 2023), it requires accurate models for the stellar spectrum and heterogeneity components, which are not currently available for M dwarfs in general although 3D models are becoming available for early-type G, K, and M dwarfs (e.g., H. N. Smitha et al. 2025). For the late-type M8 star TRAPPIST-1 in particular, the stellar spectrum exhibits features that are only partially matched by models (O. Lim et al. 2023; M. Radica et al. 2025). The M8-type stellar host also poses other challenges from an observational standpoint, including frequent stellar flares, with variable shapes and wavelength-dependent signatures, which often affect space-based observations (W. S. Howard et al. 2023).

In terms of atmosphere prospects, contrary to TRAPPIST-1 d, planets b and c were already a priori expected to be stripped of an atmosphere through hydrodynamic escape because they lie well within the runaway greenhouse limit, receiving 4.2 and 2.2 times the Earth’s insolation flux (E. Agol et al. 2021; J. Krissansen-Totton & J. J. Fortney 2022). Even assuming that TRAPPIST-1 b and c are airless, accounting for the shared irradiation history, magma ocean geochemical and thermal evolution, and plausible initial volatile budgets of the seven planets, leads to about a 70% probability of atmospheric retention for TRAPPIST-1 d (J. Krissansen-Totton 2023; M. T. Gialluca et al. 2024)—although other models are more pessimistic (G. Van Looveren et al. 2024). Whether the planet retains an atmosphere or not, however, depends on its initial volatile inventory. Atmospheric reconnaissance can therefore serve as a powerful constraint on its initial formation conditions, and is also motivated by the growing community effort (S. Redfield et al. 2024) to constrain the location of the “cosmic shoreline” (K. J. Zahnle & D. C. Catling 2017): the region of parameter space where rocky planets can retain or revive atmospheres.

Due to the larger orbital distance and cooler equilibrium temperature of TRAPPIST-1 d compared to planets b and c, transmission spectroscopy is the method of choice for atmospheric reconnaissance of this planet. TRAPPIST-1 d

has already been the target of multiple transmission spectroscopy campaigns with the Hubble Space Telescope (HST) and Spitzer Space Telescope (Spitzer; J. de Wit et al. 2018; E. Ducrot et al. 2018, 2020). So far, observations have only ruled out cloud-free, solar-metallicity, low mean molecular weight atmospheres, while not reaching the sensitivity required to search for more compact, high mean molecular weight atmospheres. With its high precision and wide wavelength coverage, the PRISM mode of JWST’s Near-Infrared Spectrograph (NIRSpec) instrument (S. M. Birkmann et al. 2022) has the potential to detect and identify potential high mean molecular weight atmospheres, as well as to distinguish the wavelength-dependent signatures of clouds or hazes from those of molecular species. At the longest wavelengths, the TLS effect is also expected to play a minor role in contaminating the planetary spectrum (S. Seager & A. I. Shapiro 2024).

Here we present the first JWST transmission spectrum of TRAPPIST-1 d, obtained with NIRSpec/PRISM from two transit observations. We first outline the observations and data reduction in Section 2, and describe the light-curve fitting in Section 3. Our methods for modeling the contributions of different stellar surface components to the transmission spectra and the out-of-transit stellar spectra are described in Section 4 and discussed in Section 5. We present our methods and results from atmosphere retrievals in Section 6. We discuss our results in Section 7 and their implications in Section 8, and conclude in Section 9.

2. Observations and Data Reduction

2.1. Summary of the Observations

JWST observed two consecutive transits of TRAPPIST-1 d with NIRSpec/PRISM on UTC 2022 November 5 and November 9. Both transits were observed as part of the Cycle 1 “NIRISS Exploration of the Atmospheric Diversity of Transiting Exoplanets (NEAT)” Guaranteed Time Observations (GTO) program GTO 1201 (PI: David Lafrenière). Each NIRSpec/PRISM bright object time series transit observation lasted 3.2 hr and consisted of 7200 integrations with six groups each, i.e., an effective integration time of 1.6 s. We chose the SUB512 (512×32 pixels) subarray, which is wide enough in the cross-dispersion direction to allow for background subtraction using non-illuminated regions away from the trace. We used the NRSRAPID readout pattern and the S1600A1 (1.6×1.6) slit. We deliberately chose to saturate the detector between $1.1\text{--}1.8\ \mu\text{m}$ (using six groups instead of the maximum of three to avoid reaching saturation) to optimize the integration efficiency and the precision reached over the longer wavelengths (70% more photons collected for each transit when factoring in the better duty cycle). This strategy enables us to increase our sensitivity to molecules that have significant opacity at $>2\ \mu\text{m}$. Both transit observations covered approximately 1.8 hr of pretransit baseline and about 35 minutes after the transit.

2.2. Data Reduction

We perform the data reduction for both time series observations using the `jwst` package, the `Eureka!` data reduction pipeline (T. Bell et al. 2022), and custom routines.

We first perform ramp fitting from the uncalibrated data products (standard Stage 1 pipeline) using the `jwst` pipeline,

and custom routines. We follow the `jwst` pipeline’s default settings, except for the few differences we outline here. We skip the jump detection step, which erroneously flags as outliers large fractions of the detector pixels, and instead rely on spatial and temporal outlier flagging at later stages in the data reduction. We perform on each image an equivalent of the reference pixel correction using the top and bottom six rows as reference pixels (with the spatial coordinate varying across rows), since no reference pixels are available on the NIRSpec detector. For each image, we calculate the median of odd and even rows among the defined reference pixels and subtract this median from all odd and even rows across the image. This reference pixel step is performed once before and once after the nonlinearity correction step.

We use a custom nonlinearity correction step, where we perform the same steps as in B. Benneke et al. (2025, submitted), where it was applied to other NIRSpec/PRISM time series observations of TRAPPIST-1. We start by using the default linearity correction coefficients on the median image, which leaves a residual curvature in the recorded flux as a function of the group number especially for highly illuminated pixels. We fit a quadratic function to this residual curvature as a function of the recorded flux in each row, which we use to correct all the images prior to ramp fitting (see B. Benneke et al. 2025, submitted, for a detailed discussion). We add a custom routine for group-level background subtraction, where we record and subtract from each group image the median of the background pixels in each detector column for that group over all integrations, after discarding 3σ temporal outliers.

The detector was deliberately saturated over a fraction of the set of six groups per integration over a limited number of pixels from approximately 1.08 and $1.77\ \mu\text{m}$ to optimize the signal-to-noise ratio (S/N) and duty cycle in the rest of the spectrum. We use the default saturation detection routine from the `jwst` pipeline, and flag as saturated not only the affected pixels but the 3×3 pixels centered on this pixel to account for charge migration. We tested a more aggressive treatment excluding from the ramp fitting not only the saturated group and subsequent groups, but also the previous group. We found that this did not impact our results and proceeded with the standard setup. Finally, we perform the ramp fitting step using the `jwst` pipeline algorithm.

We use the Stage 2 pipeline from `Eureka!` (a wrapper around the `jwst` pipeline’s Stage 2) to perform the final calibration steps prior to extraction (Stage 3). We follow the default steps, except that we skip the photometric calibration and the flat-fielding steps, which are not required for transmission spectroscopy. We skip the `extract1d` step since we perform an independent spectral extraction using `Eureka!`.

We use `Eureka!` to perform a second iteration of background subtraction as well as optimal spectral extraction on each of the integrations, to convert the slope images into a time series of observed stellar spectra. We use rows 4 to 28 in the spatial direction and columns 14 to 495 in the dispersion direction. Pixels with any flags other than saturation flags are masked. The vertical position of the trace is identified by summing the image over detector columns and fitting a Gaussian profile to the resulting flux distribution. We perform background subtraction using the same routine as for Stage 1 to remove any remaining $1/f$ noise. We rely on an additional step of background subtraction where we consider rows 7 pixels or more away from the trace, and where we discard pixels that are 10σ outliers

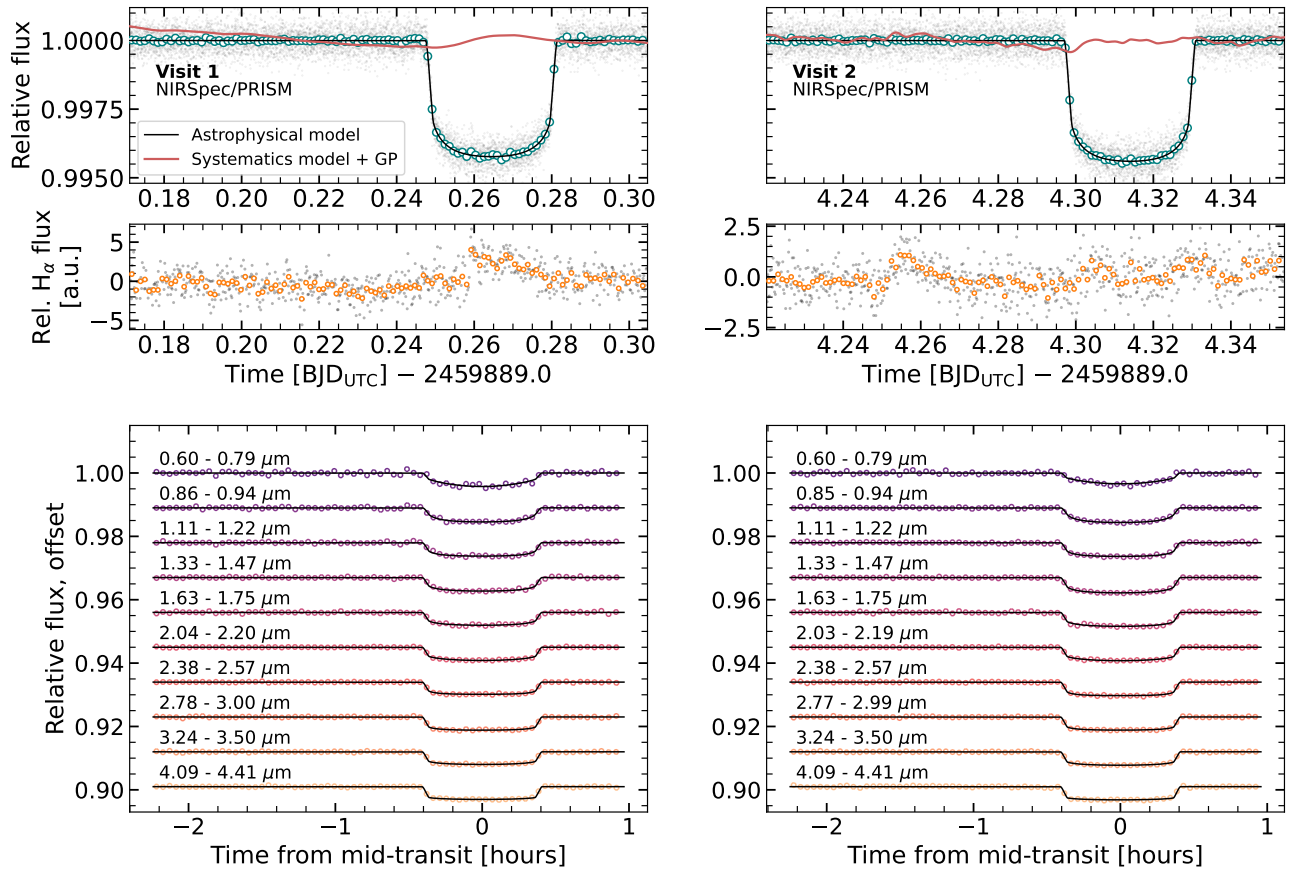


Figure 1. White and spectroscopic light-curve fitting results for the first (left) and second (right) NIRSpect/PRISM transit of TRAPPIST-1 d. Top panels: white light curves corrected by the best-fitting systematics models (gray points, and teal points are 60 point bins). The best-fit astronomical model (black) and systematics model including the Gaussian process (GP) component (red) are displayed on top of the data. Middle panels: median-subtracted integrated $H\alpha$ fluxes (in arbitrary units, multiplied by 10^4) for visits 1 and 2, binned by a factor of 10 (gray points) and 60 (orange points). The timing and temporal shape of the astrophysical systematics captured by the GP strongly correlate with the $H\alpha$ time series, suggesting that the flux variations are associated with small stellar flares during our observations. Bottom panels: 10 sample systematics-corrected spectroscopic light curves, binned by a factor of 100 (colored points). The corresponding wavelength bins are labelled, and the best-fit astrophysical model is overlaid (black line). The light curves are offset relative to each other for clarity. The first spectroscopic bin ($0.60\text{--}0.79\ \mu\text{m}$) is most affected by stellar activity and contains the $H\alpha$ line.

in time when constructing the background, in two outlier flagging iterations. Then, we loop through the odd and even rows of each group and subtract the mean value obtained from each set of rows. We then perform an optimal extraction weighted by the flux in the median image, with an aperture spanning seven rows above and below the trace position and record for each extracted spectrum in our time series the exact vertical position of the trace in the corresponding image and the trace width (standard deviation of the fitted Gaussian) in order to flag any instrumental systematics during the light-curve fitting step.

We then use EXOTEP (B. Benneke et al. 2017) to obtain the white and spectroscopic median-normalized light curves from the time series of stellar spectra. We also record potential shifts of the stellar spectrum in the dispersion direction using cross correlation with the first extracted stellar spectrum, as well as the width of each cross-correlation peak. We perform further outlier removal prior to passing on the light curves to the fitting step: we clip data points that are 3σ outliers in the spatial position of the trace on the detector, in the width of the Gaussian fitted to the trace, in the recorded spectral shifts in the dispersion direction, and in the flux (identified using a running median with a window size of 20).

We also record the integrated $H\alpha$ flux in each of our spectra, since jumps in this metric relative to its baseline have been found to track stellar flares on TRAPPIST-1 (O. Lim et al. 2023;

W. S. Howard et al. 2023). Since NIRSpect/PRISM does not have the high resolution required to resolve the $H\alpha$ line, we need to use a wider wavelength window compared to previous NIRISS/SOSS observations of TRAPPIST-1 (O. Lim et al. 2023; M. Radica et al. 2025), and measure the combined flux emitted by $H\alpha$ and a neighboring feature²⁷ (O. Lim et al. 2023). We estimate the $H\alpha$ emission by fitting a straight line to the pseudocontinuum on either side of the blended spectral feature ($0.005\ \mu\text{m}$ on each side of the $0.6460\text{--}0.6645\ \mu\text{m}$ feature). This spectrum region is then normalized, and the continuum is subtracted out before integrating the flux contained over the considered wavelength range. This metric then corresponds to the $H\alpha$ flux, scaled by a factor of 10^4 in Figure 1.

3. Transit Light-curve Fitting with EXOTEP

We fit the broadband and wavelength-dependent light curves using the EXOTEP pipeline (B. Benneke et al. 2017, 2019a; see Appendices B.1 and B.2). We evaluate the impact of our choice of prescription for limb darkening on the spectrum, and choose to adopt limb-darkening coefficients

²⁷ Here we refer to the emission feature just shortward of $H\alpha$ as seen in Figure 8(b) of O. Lim et al. (2023). Since this feature remained consistent between quiescent and flaring epochs (O. Lim et al. 2023), we conclude that the relative variations we are measuring indeed reflect the $H\alpha$ luminosity.

measured from an observation of another TRAPPIST-1 planet transit that happened during a time of relative stellar quiescence (no flares and no stellar contamination detected; see Appendix B.2.1). We examine the $H\alpha$ luminosity time series and identify small increases reflecting stellar flares (Figure 1). Such “microflares” were found to be frequent in a 5 yr HST monitoring campaign (D. Berardo et al. 2025). The flares occurred during the first transit, and in the pretransit baseline as well as during the transit in the second visit. After testing several detrending methods to account for the wavelength-dependent effect of these small flares (e.g., Figure B1), we select a systematics model that captures a linear trend with time and models the remaining correlated noise as a GP with a Matérn 3/2 kernel (see Section B.2.2).

We find that the planetary and orbital parameters inferred from the fits to visit 1 and visit 2 are in very good agreement (Figure B2, Table B1) and use them to calculate derived parameters for TRAPPIST-1 d (Table A1). The planetary radius, semimajor axis, and inclination inferred from our two NIRSpec/PRISM transits (Table A1) are all in excellent agreement with literature values obtained from fitting all the Spitzer, ground-based, HST, and K2 transit observations of the TRAPPIST-1 planets, with comparable uncertainties (E. Agol et al. 2021). Our measured transit times (Table B1) also closely match (with a precision improved by a factor of 100) the predicted times of $\text{BJD} = 2459889.265414 \pm 0.003625$ and $2459893.314955 \pm 0.003632$ for the two visits, from the dynamical forecast informed by previous transit observations of the system (E. Agol et al. 2021).

Interestingly, despite the good agreement in the inferred planetary parameters between both visits (Table B1, Figure B2), we notice significant differences in the shapes between the shapes of the two transmission spectra (Figure 2, and Table B2), which we attribute to the TLS signal varying over the span of 4 days between the first and the second observation, or approximately $1.2\times$ the stellar rotation period of 3.3 days (Table A1; R. Luger et al. 2017).

4. Stellar Surface Modeling

We leveraged the information from both the transit spectra and the out-of-transit stellar spectra to infer the distribution of dark and bright heterogeneities on the stellar surface during each visit. We assessed the visit-to-visit variability in the TLS signature, and evaluated the extent to which the independent constraints from out- and in-transit analyses painted a consistent picture of TRAPPIST-1’s surface at the epoch of the observations. We find that a quiet photosphere with a minor coverage ($\sim 10\%$) of unocculted bright heterogeneities produces the best match to our observations, while the properties of potential dark heterogeneities are unconstrained and their addition does not lead to a better fit quality.

4.1. Methodology

We obtain flux-calibrated out-of-transit stellar spectra using a modified version of the data reduction described previously. Since this is the first description of flux-calibrated NIRSpec/PRISM data to our knowledge, we describe how we customized the Eureka! pipeline for this purpose (see Appendix C.1). In order to interpret each visit’s out-of-transit flux-calibrated spectra in terms of the stellar surface properties, we present a new out-of-transit spectral retrieval submodule, Exoplanet Host

out-of-transit Spectrum Fitting Framework (exotune), implemented into the open-source stellar contamination modeling code `stctm`²⁸ (C. Piaulet-Ghorayeb et al. 2024; C. Piaulet-Ghorayeb 2024, 2025, submitted). Our procedure provides a statistically robust way to obtain posterior distributions for the relevant stellar photosphere and heterogeneity properties within a Bayesian framework rather than performing a grid search for the highest-likelihood model (e.g., H. R. Wakeford et al. 2019; S. E. Moran et al. 2023; F. Davoudi et al. 2024). In contrast to previous work (e.g., M. Radica et al. 2025), our out-of-transit stellar spectral retrievals are fully parallelized, enabling fast computation even for the large number of data points expected from spectra obtained with JWST/NIRISS/SOSS or NIRSpec. The methodology applied to infer the properties of the stellar surface is described in Appendix C.2. This cross check serves as further validation of the stellar surface heterogeneity properties we infer from the standard `stctm` fit to the impact of TLS on the transmission spectra of TRAPPIST-1 d (see Appendix C.3).

We use the results from the `stctm` retrieval to create “TLS-corrected” versions of the transmission spectra ($D_{\lambda,\text{corr}}$) for each visit. These spectra are obtained by dividing each measured transmission spectrum $D_{\lambda,\text{meas}}$ by the best-fitting model from the three-component TLS-only retrieval ($D_{\lambda,\text{TLS}}$), and multiplying the ratio of these spectra by the value of the bare-rock transit depth ($D_{\text{scale,TLS}}$; see Section C.3) corresponding to the best-fit model (3587 ppm for visit 1, 3658 ppm for visit 2):

$$D_{\lambda,\text{corr}} = \frac{D_{\lambda,\text{meas}}}{D_{\lambda,\text{TLS}}} \times D_{\text{scale,TLS}}. \quad (1)$$

The correction aims at effectively dividing the measured transmission by the stellar contamination factor $\epsilon_{\lambda,\text{het}}$, since $D_{\lambda,\text{TLS}}$ is understood in the `stctm` modeling formalism as

$$D_{\lambda,\text{TLS}} = D_{\text{scale,TLS}} \times \epsilon_{\lambda,\text{het}}, \quad (2)$$

where $\epsilon_{\lambda,\text{het}}$ is a factor representing the contamination by stellar heterogeneities (see C. Piaulet-Ghorayeb et al. 2024, Equation C1).

We leverage these spectra to perform atmosphere retrievals using the “sequential” approach, where the stellar contamination contribution is first divided out of each visit’s transmission spectrum prior to combining them for the atmosphere retrievals, and assess the impact of joint atmosphere and TLS marginalization on our sensitivity to atmosphere signatures. When displaying the combined visit 1 + 2 “TLS-corrected” transmission spectrum, we shift the visit 2 spectrum by the difference between the bare-rock transit depths of visits 1 and 2 and then show the error-weighted average of these two spectra. That spectrum is only used for visualization purposes, as we use directly the individual visit 1 and 2 TLS-corrected spectra in our SCARLET retrievals (while marginalizing over a visit-to-visit offset; see Section 6).

4.2. Inferred Stellar Surface Properties

We find that both spectra can be fully explained by stellar contamination (Figure 2), and find tentative evidence for the presence of bright surface heterogeneity regions (faculae) at the level of 1.8σ for visit 1 and 2.8σ for visit 2, while the addition of colder spots to the model is not favored (Section C.4). The TLS and out-of-transit spectral retrievals

²⁸ <https://github.com/cpiaulet/stctm>

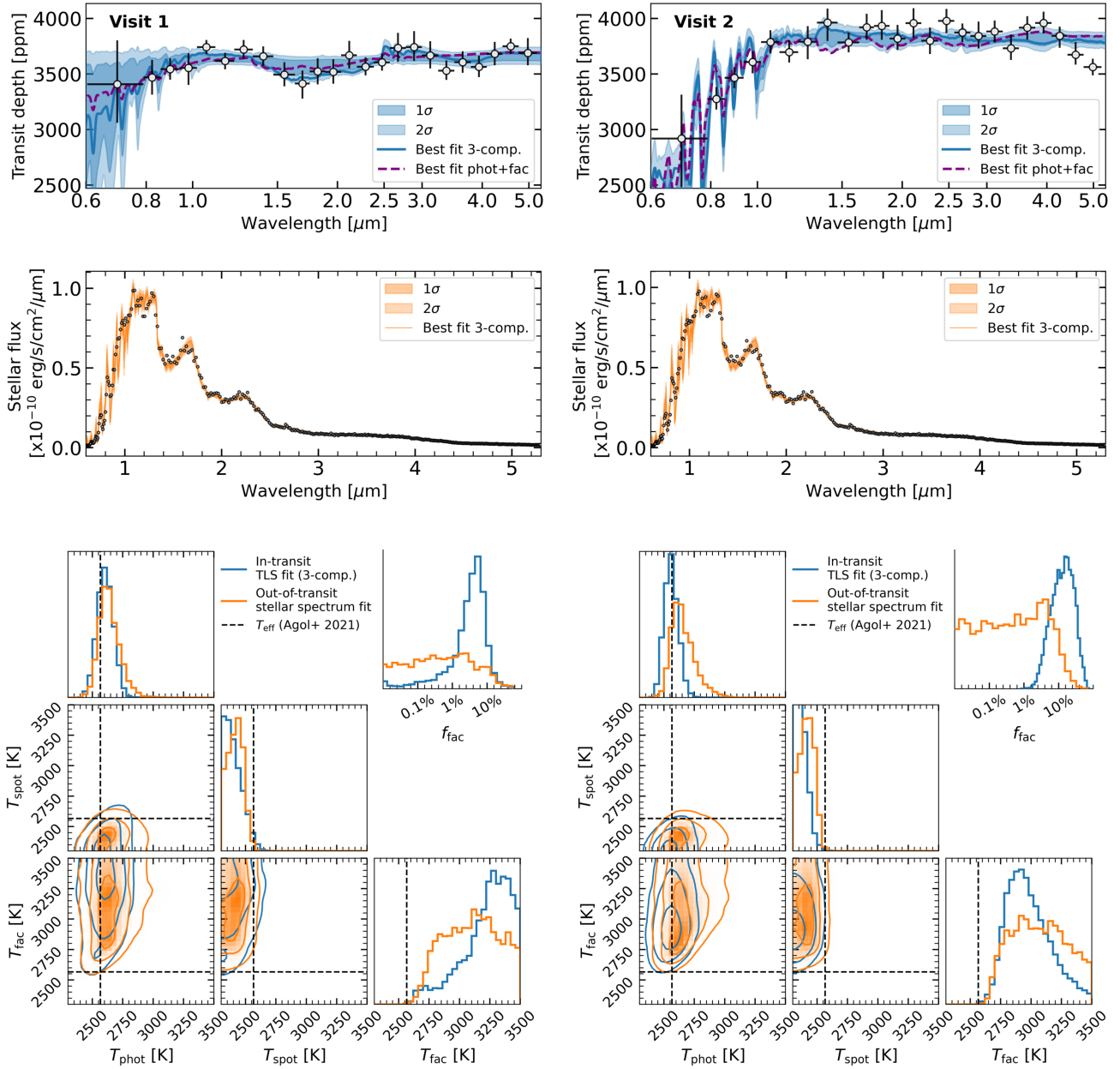


Figure 2. Results from the retrievals performed on the transmission spectra of TRAPPIST-1 d and out-of-transit stellar spectra of TRAPPIST-1 with *stctm*. The retrievals assume that the spectra can be fully explained by a three-component model including the stellar photosphere, and both cooler (“spot”) and brighter (“facula”) heterogeneity components, with no wavelength-dependent atmosphere signature in the transmission spectrum. Top panels: measured transmission spectra (black points) and best-fit models (blue) along with the 1σ and 2σ retrieved contours on the stellar spectrum (color shadings). The dashed purple line corresponds to the best-fit model in a two-component retrieval without spots. Second panels: same as for the top panels, but for the retrieval performed on the out-of-transit stellar spectra, in shades of orange. Bottom panels: joint and marginalized posterior probability distributions of the temperatures of the photosphere, spot, and facula components from both types of retrievals, with 1σ , 2σ , and 3σ contours. The independently determined stellar effective temperature from E. Agol et al. (2021) is shown with dashed lines. We display in the upper right corner the marginalized posterior distribution of the covering fraction of the facula component (the spot covering fraction was unconstrained in all our retrievals, see Table C1). We find good agreement between the stellar properties inferred from the TLS and out-of-transit spectral retrievals.

both yield very consistent posterior constraints on the retrieved stellar surface properties, despite the uniform prior adopted for T_{phot} in the exotune retrievals (Figure 2). Further, the out-of-transit exotune retrievals provide an independent constraint on the photospheric temperature (Table C1), and place a stricter upper limit of about 10% on the covering fraction of the hotter “faculae” spectral component (Table C1).

We note that model fidelity is a pervasive issue affecting the interpretation of M dwarf spectra, and our study is not immune to these biases (see, e.g., B. V. Rackham et al. 2023; TRAPPIST-1

JWST Community Initiative et al. 2024; F. Davoudi et al. 2024; F. Jahandar et al. 2025). We find that the wider wavelength coverage of NIRSpec/PRISM is helpful in this regard as it allows us to capture the Rayleigh–Jeans tail of the stellar spectral energy distribution (SED) and is less sensitive to modeling issues at the shortest wavelengths that plagued NIRISS/SOSS interpretation (O. Lim et al. 2023; M. Radica et al. 2025). We also compared our spectrum to a SPHINX, rather than a PHOENIX, model, and found that the mismatch is at least comparable to what PHOENIX models display for the nominal literature stellar

parameters (Figure C1). We discuss these caveats in more detail in Appendix C.6.

Finally, our approach does not account for the wavelength-dependent signatures of flare signals (see Figure B1), which may affect the out-of-transit and transmission spectra. Instead, here, we adopt the approach of avoiding flare epochs when constructing the out-of-transit spectrum, and retrieving the effect of the TLS effect on transmission spectra that are effectively affected by both TLS and flares (although some of the impact of flares may be taken out by the GP fitting). A more extensive investigation of this impact (O. Lim et al. 2025, in preparation) found that flares modeled under the simplifying assumption of a time-decaying blackbody hotter than the photosphere may mimic the spectral signatures of unocculted faculae, but this conclusion only holds if no GP is fitted to the associated correlated noise, while we model the wavelength-dependent flare signature with a Matérn 3/2 GP in this work. Applying a GP to the light curves should remove this impact in theory, and might even introduce a spurious spot-like *increase* in transit depths at the shortest wavelengths where the flare signal is poorly modeled by the GP, which is not detected in our case (O. Lim et al. 2025, in preparation). A blackbody function is not a perfect match to the near-infrared flare spectra of TRAPPIST-1, (W. S. Howard et al. 2023). Instead, flare spectra are expected to exhibit short- and long-wavelength components that evolve on different timescales (see, e.g., A. F. Kowalski et al. 2013). We defer a full exploration of these effects to future work and encourage the community to invest effort into understanding the impact of both the TLS effect, flares, and their wavelength-dependent signatures depending on the amplitude and timing of the flare relative to the transit in order to better assess their impact on small-planet atmosphere exploration.

5. Complementarity of In-transit and Out-of-transit Retrievals

The main conclusion from all the retrievals we performed is consistent across the in- and out-of-transit analyses: a photosphere component and a minor facula component provide a good match to the out-of-transit stellar spectrum, and the imprint of the TLS effect on the transmission spectrum (Figure 2). We find that the main strengths of out-of-transit spectral retrievals lies in their sensitivity to the temperature of the photosphere component and their independence from planetary atmospheres, while the TLS retrievals can more accurately detect spectral components with low covering fractions because of the greater fractional impact they have on the transmission spectrum.

Strengths of out-of-transit spectral retrievals. The transmission spectrum can only probe the ratios of the photosphere to heterogeneity components through the TLS effect, and therefore relies entirely on a literature prior for the temperature of the quiet photosphere T_{phot} . Meanwhile, the out-of-transit spectral retrievals are directly sensitive to T_{phot} . For both visits, our *exotune* retrievals find photosphere temperatures in perfect agreement with independent determinations in the literature (Figure 2, Tables A1 and C1). We also obtain a stricter upper limit on the contribution of the facula component to the stellar spectrum (f_{fac}) compared to TLS retrievals alone thanks to the capacity to leverage the entire spectral SED. This upper limit likely reflects an overall constraint on the range of stellar temperatures that can explain TRAPPIST-1’s NIRSpec/PRISM

spectrum; a larger contribution from a hot component would lead to noticeable changes in both spectral shape and intensity.

Strengths of TLS effect retrievals. While the *exotune* retrievals can directly probe the emission from the stellar surface, they lack spatial information of how the stellar surface outside of the transit chord looks relative to that underlying the path of the planet’s transit. In our case, we do not detect any spot or facula crossing event, suggesting that the assumption of the planet transiting over the quiet photosphere is warranted here. Despite the limited sensitivity to temperature contrasts, a distinct advantage of the TLS retrievals relative to their out-of-transit counterparts is their ability to tease out very minor stellar spectral components. This translates to a lower limit on f_{fac} from the TLS retrieval in the case of the second visit, driven by the ~ 1000 ppm transit depth variations observed at short wavelengths. Therefore, TLS retrievals can guide the setup of out-of-transit retrievals thanks to their sensitivity to smaller heterogeneity contributions, which can be degenerate (between, e.g., a large fraction of spots versus a cooler photosphere) in the *exotune* fit. In our case, the TLS retrievals motivated the exploration of at least two spectral components (photosphere + faculae) to interpret the out-of-transit spectrum.

Importantly, while both methods can yield posterior distributions of the same parameters, the fitted quantities are fundamentally different: the TLS fit probes the ratios of the spectra for the different stellar surface components (but can be biased by potential planetary atmosphere contributions), while the out-of-transit fit is sensitive to the absolute SED. This makes these approaches complementary, and the importance of accurate stellar modeling for inferences about planetary atmospheres motivates us to argue for performing this cross check for all transiting planet observations where a stellar spectrum is available. Finally, the results from the out-of-transit stellar spectral retrievals can be used to place informed priors on the TLS components of joint atmosphere–TLS retrievals. Here, for instance, both the TLS and stellar spectral retrievals support the conclusion that there is no evidence for a difference in the representative $\log g$ of the photosphere and heterogeneity components (Table C1), a conclusion we leverage in the setup of our atmosphere retrievals.

6. Atmosphere Modeling

6.1. SCARLET Atmosphere Retrievals

We use SCARLET (B. Benneke & S. Seager 2012; B. Benneke 2013, 2015; B. Benneke et al. 2019a, 2019b; S. Pelletier et al. 2021; C. Piaulet et al. 2023) to perform both atmosphere-only (Appendix D.1) and joint atmosphere–stellar contamination (Appendix D.2) retrievals on the NIRSpec/PRISM spectra of TRAPPIST-1 d. We explore for the first time the impact of the “sequential” retrievals performed in previous works on the TRAPPIST-1 planets (O. Lim et al. 2023 for TRAPPIST-1 b; B. Benneke et al. 2025, submitted for TRAPPIST-1 g) on the atmospheric inferences. In the “sequential” retrieval approach, retrievals are performed on the TLS-corrected (visit 1 + visit 2) transmission spectrum (see Section 4.1). We compare the results from this two-step process to the inferences from a joint retrieval where atmospheric properties are shared across both visits, while stellar heterogeneity parameters are fitted independently for each visit. We run three types of retrievals: (1) assuming

atmospheres with H_2/He as the background molecules, and freely fitted abundances for other species (prior favoring low mean molecular weight atmospheres); (2) assuming pure-composition atmospheres made of either CH_4 , H_2O , CO_2 , CO , or NH_3 , all of which our NIRSpec/PRISM wavelength range is sensitive to; and (3) exploring atmospheres where N_2 is mixed with an infrared-absorbing species such as CH_4 , CO_2 , or H_2O , which allows for a comparison with solar system terrestrial compositions (see details of the implementation in Appendix D). For the joint retrievals, we assume that the $\log g$ values of the heterogeneities and photosphere are identical (based on the results from the stellar surface modeling step) but allow for both visit-specific spot and facula components (see Section D.2) since spot signatures could mimic atmospheric absorption signals and are not independently ruled out by the out-of-transit spectral analysis. Finally, we assess the extent to which the limits we place on potential atmospheric thicknesses are impacted by the refraction of stellar light away from our line of sight (see Appendix D.3) for two edge-case scenarios among the cases we explore (since the atmospheric mean molecular weight correlates with the refractive index): a water-dominated (10 bar H_2O , 1 bar N_2) composition, and a pure- CO_2 (10 bar CO_2) composition.

6.2. Constraints on Potential Planetary Atmospheres

We first perform a standard well-mixed multigas atmosphere retrieval with H_2/He as the background gases, and either the “sequential” or “joint retrieval” approach to treat the TLS effect. This test allows us to assess the compatibility of our observations with different atmospheric mean molecular weights (from hydrogen-rich to metal-enriched atmospheres) across a wide range of atmospheric thicknesses (above a thick cloud deck or a rocky surface). The shape of the joint distribution on the atmospheric metallicity (Figure 3; metallicity calculated from the abundances of all the gases included: H_2 , He , H_2O , CH_4 , CO , CO_2 , N_2 , SO_2 , and NH_3) reflects the fact that our transmission spectrum does not show detectable planetary absorption features. The precision we reach enables us to discard with a high level of confidence cloud-free hydrogen-dominated atmospheres (see also Figure 4), in agreement with previous studies (J. de Wit et al. 2018; E. Ducrot et al. 2018, 2020). However, for conditions rich in high mean molecular weight volatiles, thick cloud-free atmospheres remain compatible with the spectrum. This conclusion is likely driven by our inclusion of NH_3 , SO_2 , and CO_2 in the model (Figure 4, see Section 6.2).

This motivates us to further explore the potential chemical makeup of high mean molecular weight atmospheres compatible with the observed spectrum, and our sensitivity to the presence of specific infrared absorbers. There, we turn to the results of our single-molecule retrievals, where we fit for effective surface pressures of an atmosphere made of 100% H_2O , CH_4 , CO , SO_2 , CO_2 , or NH_3 . None of the molecules that we probed were detected (Figure 4) and we find that even thin, high mean molecular weight atmospheres of 1 bar are excluded with better than 95% confidence for 100% CH_4 , CO , H_2O , and CO_2 atmospheres in the most conservative case of a joint retrieval for the TLS and atmosphere contributions. For NH_3 , however, we only obtain a meaningful upper limit in the sequential fit (Table D1, Figure 4) which vanishes when we marginalize over all possible TLS realizations. Finally, SO_2 and CO_2 only have a few very small amplitude features within reach of NIRSpec/PRISM, especially in these conservative

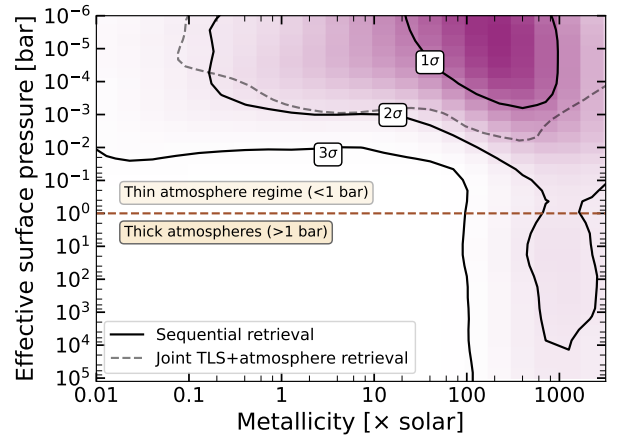


Figure 3. Constraints on the effective surface pressure of a potential atmosphere (with H_2/He assumed to be the background “filler gas”) as a function of atmospheric metallicity. We show the joint posterior probability distribution on the effective surface pressure (of either a bare-rock surface or a gray cloud deck) and the atmospheric metallicity calculated from the posterior distributions of the H_2/He , H_2O , CO , CO_2 , CH_4 , N_2 , NH_3 , and SO_2 abundances. The purple shading indicates high-probability regions, and the contour lines trace the 68%, 95%, and 99.7% confidence regions for the sequential retrieval (solid) and the 95% contour in the joint retrieval case (dashed). At the 3σ level, our spectrum excludes hydrogen-dominated atmospheres down to 0.01 bar effective pressures, and up to $100 \times$ solar metallicity, while some high mean molecular weight atmospheres remain compatible with the spectrum.

100% composition cases, which lead to the overall highest atmospheric mean molecular weights of all the tests we ran. For these two molecules, our results depend on the assumption on the TLS treatment, and are more sensitive to the details of how the stellar contamination correction is performed. This difference is likely driven by the single-molecule modeling choice, since these two species only have a few localized absorption bands over the wavelength range of NIRSpec/PRISM.

We also run retrievals under the assumption that a large part of the atmospheric mass may be contained in a high mean molecular weight species without significant infrared absorption features, such as N_2 (as is the case on Earth), which only has a weak feature from collision-induced absorption at $\sim 2.2 \mu\text{m}$. For these retrievals (focusing on the results obtained from the joint TLS + atmosphere fits), we obtain joint constraints on the partial pressures of an atmosphere made of N_2 and one of the infrared absorbers CH_4 , H_2O , or CO_2 , allowing us to compare our constraints with the compositions of solar system terrestrial atmospheres (Figure 5). We obtain upper limits for all three molecules, reaching the sensitivity required to exclude abundances greater than 500 ppb for CH_4 , 6000 ppb for CO_2 , and 1412 ppm for H_2O in a 1 bar N_2 -background atmosphere (or 8.9 ppm, 158 ppm, and 0.1%, respectively, for a thin 0.1 bar N_2 -background atmosphere) within the 95% confidence interval.

Although atmospheric refraction is crucial to consider for TRAPPIST-1 d since it orbits a small late M dwarf, we find that it does not impact the upper limits we obtain from the retrievals. Specifically, we find that in the thick 10 bar pure- CO_2 (water-rich) scenario, refraction cuts off the contribution from layers deeper than 6.61 (1.67) bar. For the CO_2 atmosphere case, this corresponds to pressures higher than the upper limits we place (Figure 4). For the N_2 - H_2O case, this cutoff only affects the deep, 10 bar H_2O + 1 bar N_2

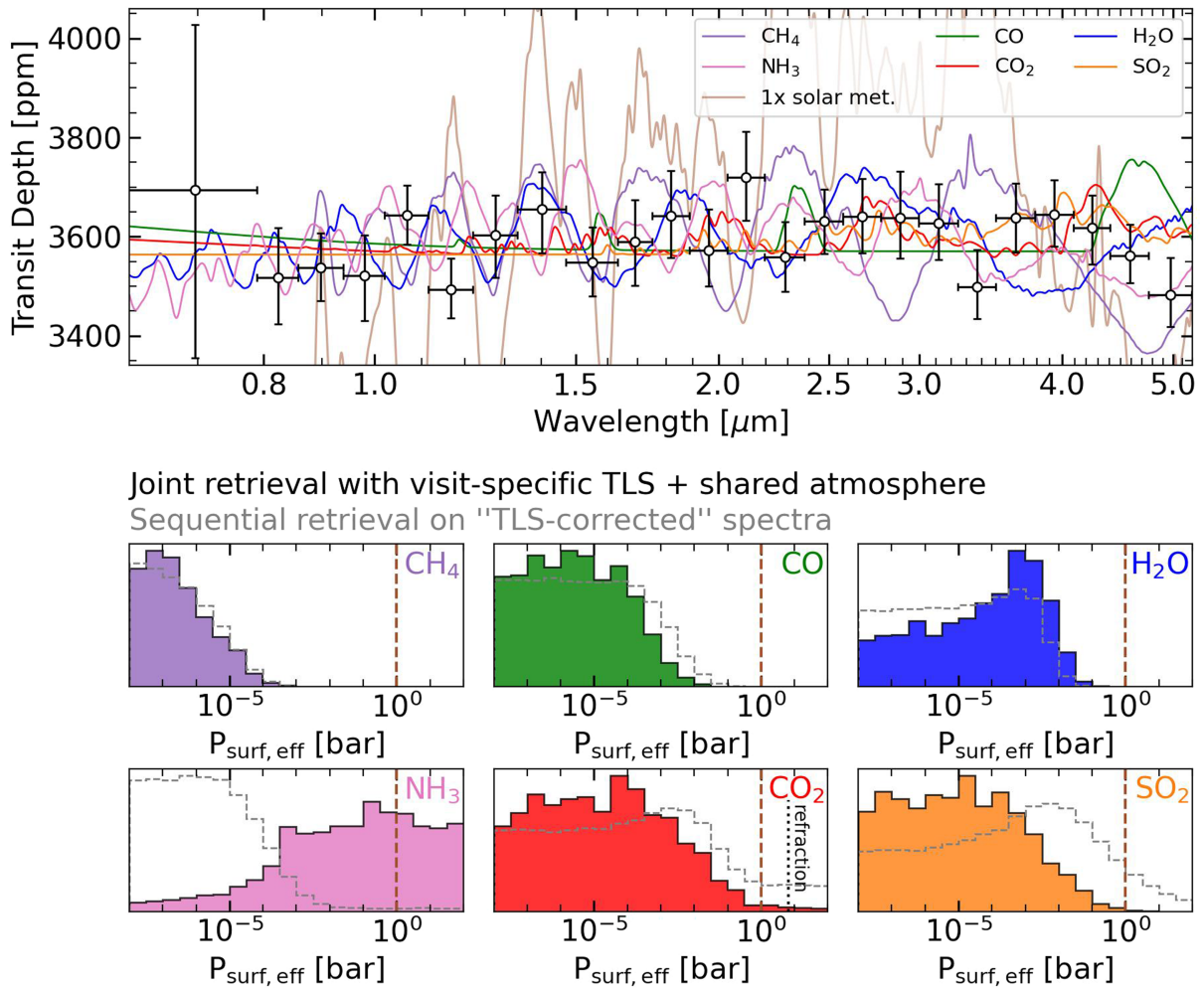


Figure 4. Constraints on the effective surface pressure of a potential high mean molecular weight atmosphere on TRAPPIST-1 d. Top panel: the combined visit 1 + 2 transmission spectrum is shown after stellar contamination correction (black points, spectrum used for the sequential retrievals). We show representative forward models for a cloud-free solar-composition atmosphere and 1 bar atmospheres made of 100% CH₄, CO, H₂O, NH₃, CO₂, or SO₂, smoothed to a resolving power of $R = 100$. Bottom panels: marginalized posterior distributions of the effective surface pressure of CH₄, CO, H₂O, NH₃, CO₂, and SO₂ (gray: sequential fit; solid filled distributions: joint fit with stellar contamination). The dashed vertical lines highlight the location of a 1 bar atmosphere. For the 100% CO₂ case, we show as a dotted line the refraction limit for a 10 bar atmosphere.

case (where the signal from pressures deeper than ~ 2 bar would be cut off), but the thinner atmosphere scenarios where most of our posterior lies are largely unaffected by this effect (Figure 5). Altogether, these results underscore the sensitivity of our TRAPPIST-1 d spectrum to high mean molecular weight atmospheres with surface pressures even thinner than on Earth, enabling detailed comparisons to solar system terrestrial atmospheres (see Section 7.2).

6.3. Impact of Stellar Contamination on the Inferences

Our modeling setup enables us to assess the impact of the retrieval assumptions on the sensitivity of our high-precision transmission spectrum to a variety of high mean molecular weight compositions. This modeling exercise opens up the possibility to uniquely assess the impact of the unknown stellar contamination on the inferences that can be made about the atmosphere, which matters substantially for TRAPPIST-1 d as both visits are seemingly affected by stellar contamination to significantly different degrees (Figure 2).

We find that, for TRAPPIST-1 d, the limits we obtain on CH₄ and CO are not substantially altered by the joint fitting of stellar

and planetary contributions (Figure 4). However, constraints on the presence of thin (< 1 bar) atmospheres are weakened for the pure-H₂O case (with an upper limit moving to pressures a factor of 2 higher), and essentially vanish for the pure-NH₃ case when visit-dependent stellar contamination is marginalized over. This behavior is largely expected for water, as we demonstrated with the same retrieval framework in M. Radica et al. (2025) for the case of TRAPPIST-1 c. This is in line with the theory of the TLS effect (e.g., B. V. Rackham et al. 2018), as specifically unocculted bright regions can imprint inverted water features on the spectrum that may mask planetary absorption signatures. Stellar contamination features, however, do not typically create NH₃-like signatures. For ammonia, our results can be traced back to the relatively smaller amplitude of NH₃ features compared to H₂O and CH₄ from the molecular cross sections (for similarly high molecular masses), and to the fact that the stellar contamination model has enough flexibility to alter the shape of each visit's transmission spectrum in a way that does not rule out NH₃ features in a joint retrieval setup. The cases of SO₂ and CO₂ are trickier to explain to the extent that the theoretically more conservative joint retrieval setup yields

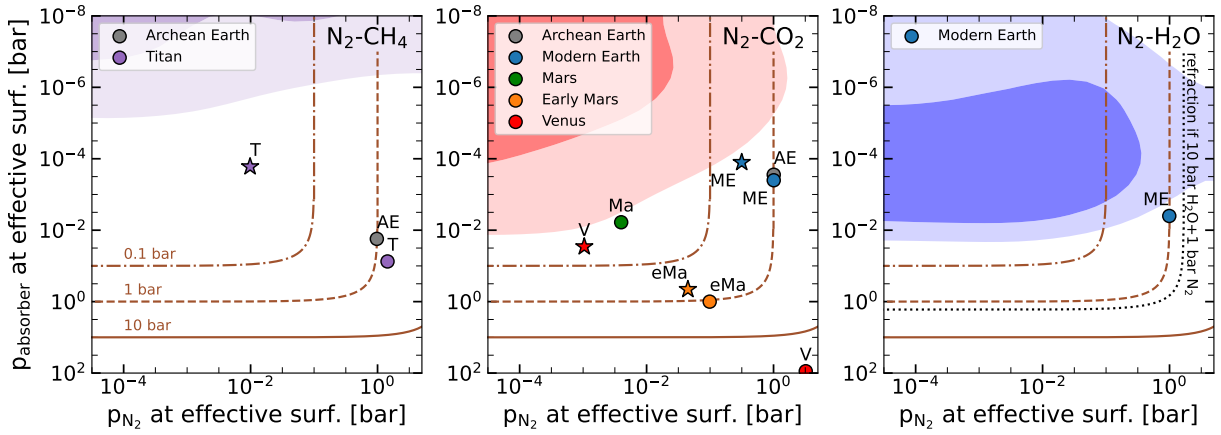


Figure 5. Joint posterior constraints on the surface partial pressure of key atmospheric absorbers for atmospheric scenarios similar to solar system terrestrials, with N_2 as the background gas. The color shadings correspond to the posterior probability mass, and the contours highlight the 68% and 95% confidence regions from the joint retrieval in each case. The absorber in the N_2 -background atmosphere is labeled in the top right corner of each panel. The dashed-dotted, dashed, and solid sienna lines correspond to an atmospheric surface pressure of 0.1, 1, and 10 bar. We illustrate where typical models of Archean Earth (“AE”), modern Earth (“ME”), and compositions for Titan (“T”), early Mars (“eMa”), Mars (“Ma”), and Venus (“V”; T. Encarnaz & A. Coustenis 2018) approximately lie relative to our constraints for TRAPPIST-1 d. Star markers correspond to specific cloudy cases, where the planet is modeled as either cloudy or cloud-free (for Titan, modern Earth, and Venus). For the N_2 - H_2O case, we show as a black dotted line the limit of pressures that would be cut off for a case that is even thicker than the upper bound in our model (10 bar H_2O + 1 bar N_2). For thinner atmospheres, the impact of refraction on atmosphere detectability is even weaker.

stricter constraints than the sequential retrieval. It can, however, be understood in the context that the CO_2 and SO_2 features are much more localized than their CH_4 , H_2O , or NH_3 counterparts: in the 100% composition case, the model spectrum is essentially flat throughout most of the NIRSpec/PRISM wavelength range with a few absorption features that can be cut off by clouds. If these happen to line up with small deviations in the TLS-corrected spectra (Figure 4), the constraints may be wider in the sequential case than in the joint retrieval case, where the parameters of the TLS contribution are optimized to yield a flat spectrum at least up to about $1.5 \mu m$. This explanation is in line with the similarity of the results we obtain for CO (to a lesser extent), which also has only a few features between 0.6 and $5.2 \mu m$. Still, none of the compositions allowed by the posterior probability distribution from the sequential fit case are formally significantly ruled out in the joint fit setup, which is consistent with the sequential fit corresponding to a smaller, but nested, prior probability space.

7. Discussion of the Remaining Atmosphere Prospects

Our precise transmission spectrum can be fully explained by stellar contamination alone, and therefore enables us to rule out cloud-free or thick atmosphere scenarios across a wide range of potential atmospheric metallicities (Figure 3). We explore the plausibility of the atmospheric scenarios that cannot be ruled out from the transmission spectrum alone: a thin or cloudy hydrogen-rich atmosphere, or a high mean molecular weight atmosphere. We discuss our sensitivity to atmospheres similar to solar system terrestrials, the impact of 3D atmosphere circulation and cloud condensation on the detectability of a water-rich atmosphere, as well as the implications of a bare-rock scenario for the formation of TRAPPIST-1 d and for atmosphere prospects on the colder planets TRAPPIST-1 e, f, and g.

7.1. Ruling Out Thick Hydrogen-rich Atmospheres

The joint effective surface pressure–atmospheric metallicity posterior obtained for TRAPPIST-1 d (Figure 3) leaves only

two possibilities for H_2 -rich, low-metallicity atmospheres: either we are seeing all the way down to the planetary surface, and TRAPPIST-1 d has an extremely thin (less than 0.01 bar at the 3σ level) hydrogen-rich atmosphere, or the conditions are met for clouds or hazes to form at pressures of $\lesssim 0.01$ bar.

The cloudy case is disfavored by both modeling and laboratory work. Even if atmosphere observations alone may not rule out cloudy hydrogen-rich atmospheres, high-altitude clouds or hazes are extremely unlikely to form at low pressures, muting spectral features for TRAPPIST-1 d (C. V. Morley et al. 2015; S. E. Moran et al. 2018). More generally for H_2 -rich atmospheres, we have to consider both sources and sinks of the atmospheric H_2 . From a formation standpoint, only small amounts of H_2 could have been accreted by TRAPPIST-1 d from the protoplanetary disk during the formation of the rapidly migrating planet chain (Y. Hori & M. Ogihara 2020). Further, any primordial hydrogen was likely lost over time due to intense X-ray–ultraviolet (XUV) irradiation from the host star during its lengthy pre-main-sequence phase, and continued atmospheric loss over the system’s lifetime (Y. Hori & M. Ogihara 2020; M. Turbet et al. 2020). Although H_2 can also be provided by outgassing from a mantle with a reduced composition, modeling of the outgassing rate found it to be too low to replenish a hydrogen-dominated atmosphere relative to the escape rate (R. Hu et al. 2023).

Therefore, we conclude that (1) thick cloud-free hydrogen-rich atmospheres are ruled out by our transmission spectrum; (2) thin H_2 -rich alternatives are strongly disfavored when considering TRAPPIST-1 d in the context of its formation and evolution under stellar irradiation; and (3) high-altitude clouds or hazes are not expected to form on TRAPPIST-1 d if it has a low-metallicity atmosphere.

7.2. Plausibility of a High Mean Molecular Weight Atmosphere

We find that atmospheres thicker than 1 bar dominated by CH_4 , CO , or H_2O are excluded with high confidence (Figures 4 and 5, Table D1). The absence of CH_4 is not surprising, as

methane is expected to be quickly depleted under the irradiation conditions of TRAPPIST-1 d, from its photodissociation by incoming stellar photons and the formation of organic hazes (M. Turbet et al. 2018, 2020), except in the presence of a very strong source, e.g., biotic methanogenesis (J. Krissansen-Totton et al. 2018; V. S. Meadows et al. 2023). However, CO is a potential outcome for outgassed atmospheres from mantle-atmosphere geochemistry (e.g., P. Liggins et al. 2021), although its abundance can vary widely depending on the redox state of the mantle, and the mantle melt fraction (D. J. Bower et al. 2022).

The constraints on pure-H₂O atmospheres are more nuanced. The upper limit on the effective surface pressure for H₂O is similarly constraining to those of CH₄ and CO atmospheres but may reflect the formation of high-altitude water clouds rather than a thin H₂O atmosphere (see next subsection).

Although marginally disfavored by the transmission spectrum, atmospheres dominated by NH₃, CO₂, or SO₂ have lower-amplitude features that cannot be completely ruled out given our error bars (Figure 4). However, NH₃ is not expected to be a stable atmospheric component, as it is rapidly depleted by photodissociation, similarly to methane (M. Turbet et al. 2018). Further, interior degassing can produce SO₂, but in lower amounts than CO₂, making the SO₂-only atmosphere an unlikely edge case (P. Liggins et al. 2021; D. J. Bower et al. 2022). A mixed CO₂-SO₂ atmosphere, indicative of a highly oxidized mantle, remains within the realm of possibility—although we note that the majority of the posterior mass for CO₂ still lies at very low effective surface pressures.

We then turn to comparing the constraints we obtain on high mean molecular weight atmospheres where an infrared absorber is mixed with N₂ to constrain further the plausible composition of an atmosphere on TRAPPIST-1 d (Figure 5). Comparing with the atmospheres of solar system terrestrial bodies, we find that our methane constraints enable us to confidently exclude Titan-like atmospheric compositions (B. Charnay et al. 2014) with 1.6% CH₄ in an N₂-background atmosphere at better than 3 σ . Even when considering the “cutoff” effect of Titan’s haze layer on the transmission spectrum at 0.01 bar (“cloudy Titan” case; Figure 5) and the lower CH₄ abundance measured above Titan’s haze, Titan-like conditions can be ruled out at better than 95% confidence. Methane was also present on Earth during the Archean eon, which we model following previous calculations for TRAPPIST-1 d (V. S. Meadows et al. 2023). We find that the amounts of CH₄ in a 1 bar Archean Earth atmosphere are incompatible with our observations at the 3 σ level. This result holds even when considering the entire range of methane partial pressures from other models of Earth’s atmosphere during this eon (approximately 10⁻⁵–10⁻² bar; D. C. Catling & K. J. Zahnle 2020).

We similarly compare our results in the N₂-CO₂ and N₂-H₂O cases to the composition of modern Earth (as defined in V. S. Meadows et al. 2017; atmospheric water abundance on results from Y. Kasai et al. 2011; V. Kelsey et al. 2022). We also show where a clear or cloudy Venus, a clear or cloudy early Mars, or a thin-atmosphere modern Mars would lie in the N₂-CO₂ parameter space. The cloudy Venus-like composition assumes a 0.03 bar cloud deck in a thick CO₂-rich atmosphere. For early Mars, we adopt a 1.1 bar surface pressure including 1 bar of CO₂ (R. Hu & T. B. Thomas 2022). The cloudy early Mars case includes CO₂ ice clouds condensing out at ~0.5 bar, lowering the effective surface pressure (M. Turbet & F. Forget

2021). For modern Mars, we adopt a 0.01 bar surface pressure. We find that we can exclude very thick CO₂ atmospheres, such as a clear Venus composition, as well as the amounts of CO₂ expected for the Archean Earth (similar in CO₂ content to the pre-industrial Earth scenario modeled by V. S. Meadows et al. 2023). Although the CO₂ partial pressure in the atmosphere of Archean Earth could range from 10⁻³ to 1 bar (D. C. Catling & K. J. Zahnle 2020), the value displayed in the figure corresponds to the smallest realistic amounts of CO₂ in such a scenario and this model uncertainty does not impact our conclusions. Finally, we can rule out the early Mars scenario with better than 95% confidence (even when accounting for the CO₂ ice cloud presence)—although here also, the exact composition of Mars is uncertain and may have varied through the ages (E. S. Kite 2019). Only extremely thin or cloudy atmospheres, such as a 0.01 bar Mars-like atmosphere very rich in CO₂, or cloudy cases similar to Venus’ sulfuric acid or Earth’s water clouds remain marginally consistent with our observations (Figure 5).

7.3. Implications of a “Hot Start” Scenario on the Detectability of a Water-rich Atmosphere

The lack of observed atmospheric absorption features in the transmission spectrum of TRAPPIST-1 d (Figure 4) cannot conclusively rule out the presence of a high mean molecular weight atmosphere, especially if the conditions are met for aerosols to form in the upper atmosphere at the terminator (Figure 3). We turn to 3D GCM simulations in order to evaluate the possibility that TRAPPIST-1 d’s transmission spectrum can be explained by the formation of such aerosols in a water-rich atmosphere.

The traditional suite of GCMs used to define habitable conditions (i.e., liquid water at the planetary surface) for rocky planets in synchronous rotation around M dwarfs (J. Yang et al. 2013; R. K. Kopparapu et al. 2016; R. Kopparapu et al. 2017) assume “cold start” conditions (the water is initially condensed out) to delineate the inner edge of the HZ as the lowest instellation required to trigger the loss of the water ocean through the runaway greenhouse effect. In such scenarios, the albedo feedback provided by low-altitude dayside clouds can be sufficient for TRAPPIST-1 d to have habitable conditions. However, it is likely that planets instead start “hot” (M. Ikoma et al. 2018; T. Lichtenberg & Y. Miguel 2025): the gravitational potential energy released after formation, as well as the potential greenhouse effect of any accreted volatiles, can bring the interior of a planet to temperatures high enough to reach a completely molten state. The mantle of TRAPPIST-1 d could then remain at least partially molten over geological timescales as it gradually crystallizes (e.g., A. C. Barr et al. 2018; A. Grayver et al. 2022).

In such a “hot start” scenario, we show that if abundant water is present in the atmosphere, cloud formation is predicted by GCM modeling. Further, TRAPPIST-1 d’s insolation places it in a regime where such clouds could essentially preclude the detection of water-rich atmospheres using transmission spectroscopy. Indeed, 3D GCM modeling predicts that the present-day orbit of TRAPPIST-1 d (currently receiving a flux of 1.125 S_{\oplus} ; Table A1) would be interior to TRAPPIST-1’s water condensation limit (at about 0.772 S_{\oplus}), which is the limit in insolation flux where the planet would cool down enough over its lifetime to condense a surface liquid water ocean. Therefore, any water initially accreted by

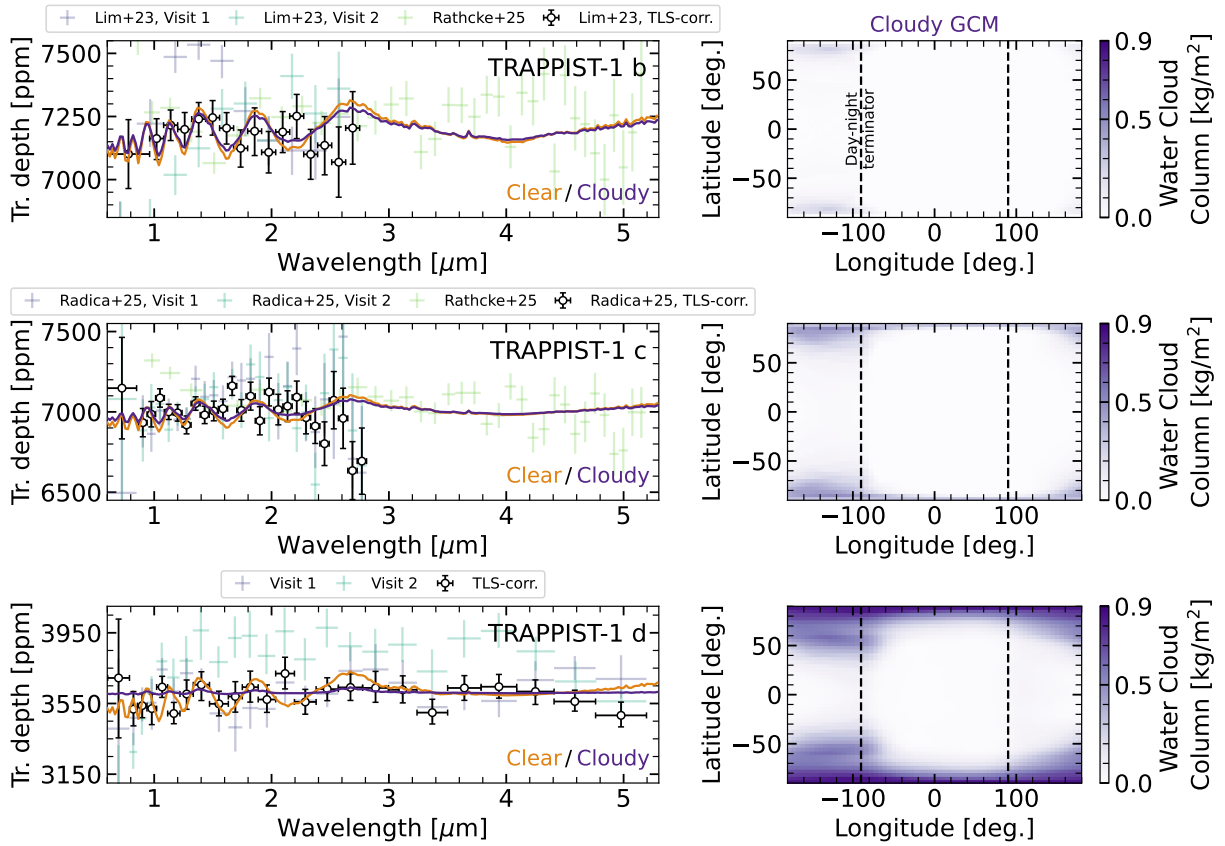


Figure 6. Comparison of the transmission spectra of TRAPPIST-1 b, c, and d, with 3D GCM simulations. The left panels show the measured NIRISS/SOSS transmission spectra of TRAPPIST-1 b and c (O. Lim et al. 2023; M. Radica et al. 2025), the NIRSpec/PRISM transmission spectra of TRAPPIST-1 b and c (A. D. Rathcke et al. 2025), as well as the NIRSpec/PRISM transmission spectra of TRAPPIST-1 d obtained in this work (colored points). Binned versions of the spectra for which `stctm` retrievals were published, after correction by the best-fitting TLS-only model (offset by 450 ppm for TRAPPIST-1 c) are shown as the black points. Binning was applied to the NIRISS/SOSS order-one spectra of TRAPPIST-1 b (five points together), to the TLS-corrected spectrum of TRAPPIST-1 b (three points together), and to the NIRSpec/PRISM spectra of TRAPPIST-1 b and c (10 points together), for visualization purposes. We show simulated spectra from 3D GCM simulations of a water-rich composition (10 bar H_2O + 1 bar N_2) on each planet, for cloudy (purple) and cloud-free (orange) conditions. While clouds barely affect the GCM model spectra in the cases of TRAPPIST-1 b and c, they reach the terminator and flatten the transmission spectrum at the insolation of TRAPPIST-1 d. The right panels illustrate this by showing the water cloud columns as a function of latitude and longitude for each cloudy GCM simulation (shades of purple, terminator shown with black dashed lines).

TRAPPIST-1 d or outgassed by mantle geochemistry should remain in the atmosphere (M. Turbet et al. 2023). However, contrary to the inner planets TRAPPIST-1 b and c, given its lower insolation, \sim millibar-level stratospheric water clouds form on the nightside of a water-rich N_2 – H_2O atmosphere on TRAPPIST-1 d and reach up to the terminator region (Figure 6). Such high-altitude clouds would provide a plausible explanation for our observed flat transmission spectrum, given the current level of precision (Figure 6). These results are seemingly contrasting with previous work that found that the presence of such stratospheric clouds in an Earth-like atmosphere does not impact atmospheric signatures in transmission (D. Doshi et al. 2022). However, the Earth-like stratospheric clouds are modeled at 600 to 100 mbar by D. Doshi et al. (2022), which are 2 to 3 orders of magnitude deeper than the water clouds predicted to form on the TRAPPIST-1 planets (M. Turbet et al. 2023), highlighting the importance of full planet-specific GCM simulations for tidally locked rocky planets. In comparison, for CO_2 -dominated atmospheres, the spectroscopic impact of CO_2 cloud condensation is very minimal for TRAPPIST-1 d (M. Turbet et al. 2023).

Finally, while H_2O clouds are stable across a wide variety of surface compositions, the atmosphere on TRAPPIST-1 d could

also be obscured by other cloud species expected at these temperatures. Specifically, nitrogen-bearing condensates such as ammonia (NH_3), ammonium chloride (NH_4Cl), or ammonium hydrosulfide (NH_4SH), as well as solid disulfide ($\text{S}_2[\text{s}]$), are expected for temperatures lower than ~ 400 K from crust-atmosphere equilibrium calculations (O. Herbolt et al. 2022). A volcanically active hot surface could support the condensation of graphite or soot-like particles ($\text{C}[\text{s}]$), while a thick atmosphere with a surface pressure of ~ 100 bar would allow for the condensation of solid sulfuric acid ($\text{H}_2\text{SO}_4[\text{s}]$) Venus-like clouds (M. A. Bullock & D. H. Grinspoon 2013).

8. Implications of Complete Atmosphere Stripping on Formation Conditions

Considering the planet in its irradiation context, may cast further doubt as to the likelihood of even a cloudy H_2O -dominated atmosphere. In particular, 3D GCM simulations suggest that any water on TRAPPIST-1 d would lie in its atmosphere in vapor (or condensed cloud) form, rather than in a liquid surface water ocean (M. Turbet et al. 2023). This distribution of the water inventory makes the planet more susceptible to water loss at its present-day orbital distance than

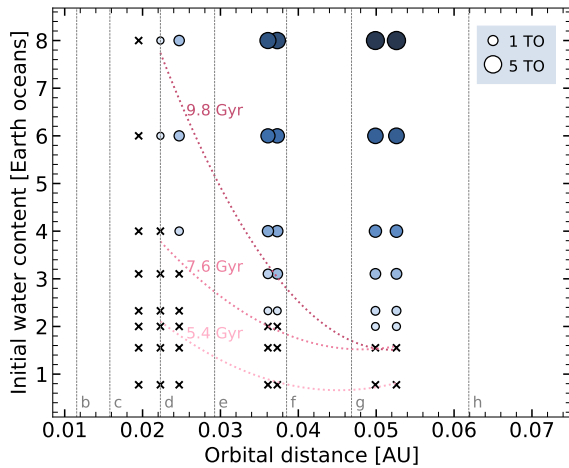


Figure 7. Results from the planetary evolution models for TRAPPIST-1 d as a function of initial water content and planetary orbital distance (following K. Moore et al. 2024). The present-day orbital distances of the seven TRAPPIST-1 planets are indicated with gray dashed lines, and labeled at the bottom. Simulation results for planets evolved for 7.6 Gyr (median age of TRAPPIST-1) are represented by circles (size and color proportional to the final water content in Earth oceans; “TO”) or by an “x” if the planet is completely desiccated at the end of the simulation. Pink lines show fits to the transition to desiccated planets when the simulation is run for 5.4, 7.6, and 9.8 Gyr. TRAPPIST-1 d could have retained water in its interior and atmosphere over 5–10 Gyr if it migrated inward only recently, or even in situ if it initially accreted an initial water budget of more than ~ 4 Earth oceans.

the cooler HZ planets TRAPPIST-1 e, f, and g, which likely exited their initial runaway greenhouse phase. The complete stripping of the atmosphere of TRAPPIST-1 d is consistent with our observed transmission spectrum, and we consider the implications of such a bare-rock planet in terms of its formation conditions and of the likelihood of present-day atmospheres on the outer TRAPPIST-1 planets.

Numerous theoretical works have found that the TRAPPIST-1 planets are especially vulnerable to the loss of atmospheric volatiles due to the long pre-main-sequence period where the stellar XUV irradiation is particularly intense (E. Bolmont et al. 2017; K. Moore et al. 2023; G. Van Looveren et al. 2024). This fragility against potential loss means that atmospheric retention is limited by each planet’s initial volatile inventory, which would have to be large enough to prevent complete stripping or allow for significant atmospheric outgassing.

Stripped atmospheres with surface pressures less than 1 bar are the most likely outcomes for TRAPPIST-1 b and c, even for initial water inventories of up to ~ 100 Earth oceans (close to the theoretical upper limit for the inner planets TRAPPIST-1 b, c, and d; J. Krissansen-Totton & J. J. Fortney 2022; J. Krissansen-Totton 2023). However, if TRAPPIST-1 d initially had a similar amount of volatiles to TRAPPIST-1 b and c, its range of plausible present-day atmospheres, even under the impact of the same host star irradiation, could a priori allow for 1–10 bar atmospheres (J. Krissansen-Totton & J. J. Fortney 2022). The nondetection of such a thick atmosphere of TRAPPIST-1 d would therefore suggest the inner planets of the TRAPPIST-1 system started out much dryer than initially believed.

We use a box model prescription (see Appendix E; K. Moore et al. 2023, 2024) to estimate the vulnerability of various initial volatile inventories to stripping by stellar

irradiation. TRAPPIST-1 d, even at its present-day orbital distance, would likely have been shielded from complete water loss if it formed with more than about four Earth oceans of water (Figure 7). Such an initial inventory corresponds to a water mass fraction on the order of 0.2%–0.3% at formation, which is consistent with estimates of Earth’s water mass fraction ranging from about 0.02% to 0.2% (for 1–10 Earth oceans) depending on how much water is trapped by minerals in the mantle transition zone (B. Schmandt et al. 2014).

Our evolution models therefore support a dry formation for the inner TRAPPIST-1 planets in the case of complete stripping of TRAPPIST-1 d’s atmosphere. TRAPPIST-1 d retains a magma ocean over ~ 500 Myr in our simulations in which water can be efficiently dissolved, essentially protecting the water inventory from escape throughout the very early and active stages of the star’s evolution. However, our prescription for the water saturation limit of the magma ocean implies that if more water is initially present than can efficiently be dissolved in the mantle, atmospheric erosion can begin right from the start of the simulation. Alternatively, TRAPPIST-1 d may have migrated inward only recently, and could therefore have been shielded from desiccation by exiting the runaway greenhouse phase earlier in its evolution.

Compared to TRAPPIST-1 d, the outer TRAPPIST-1 planets are able to more effectively shield water from loss to space in surface liquid ocean reservoirs (M. Turbet et al. 2018). This conclusion is further supported by our evolution models run for TRAPPIST-1 e, f, and g. All three planets lie outside the runaway greenhouse limit and are, respectively, 1.8, 2.7, and 3.4 times more massive than TRAPPIST-1 d; they would therefore have retained significant water inventories over ~ 5 –10 Gyr provided they formed with at least a few Earth oceans of water (Figure E1), even without invoking any migration. Further, the outer TRAPPIST-1 planets could have much larger initial volatile budgets than the inner three (J. Krissansen-Totton & J. J. Fortney 2022; J. Krissansen-Totton 2023), suggesting that even if TRAPPIST-1 d is stripped, less-irradiated planets likely still bear atmospheres provided they accreted a moderate amount of water during or after their formation.

9. Summary and Conclusions

We present the first 0.6–5.2 μm JWST transmission spectrum of the small temperate terrestrial planet TRAPPIST-1 d, obtained from two transit observations with NIRSpec/PRISM. We detect small stellar flares in both visits and find that stellar contamination has an important impact on our observations, leading to 500–1000 ppm signatures. Stellar contamination alone provides a satisfactory fit to our observations, without the need for any planetary atmosphere.

The small flares we flag are apparent in the $\text{H}\alpha$ time series, while they are otherwise barely noticeable in the white light curves, stressing the importance of monitoring spectral lines affected by flares and the potential of even the lower resolution of NIRSpec/PRISM, compared to NIRISS/SOSS, for doing so. Although small, these flares have wavelength-dependent properties and require independent detrending for each spectroscopic light curve to obtain robust transit spectra.

We find that stellar contamination is the dominant contributor to our transmission spectra, and that it varies from one visit to the next although they are only 4 days, or 1.2 stellar rotations, apart. We perform additional retrievals on the

out-of-transit stellar spectrum, in order to compare the TLS-effect-inferred stellar surface heterogeneity distributions that impact joint TLS-atmosphere retrievals to prior expectations based on the baseline stellar spectrum. We introduce *exotune*, a subroutine of the open-source *stctm* module, with capabilities to perform such retrievals, and suggest this method as a systematic cross check as it provides complementary information to standard TLS retrievals. Out-of-transit spectral retrievals provide independent constraints on the photosphere temperature to be compared to inferences based on high-resolution spectra, which can be biased by, e.g., incomplete line lists for M dwarfs (F. Jahandar et al. 2025). Improvements in model fidelity, for example using data-driven approaches (B. V. Rackham et al. 2023; TRAPPIST-1 JWST Community Initiative et al. 2024) will be crucial to provide better stellar contamination corrections and to obtain even more precise limits on potential atmospheres on the TRAPPIST-1 planets.

Once the TLS effect is accounted for, our NIRSpec/PRISM observations reach a high sensitivity even to low scale height, high mean molecular weight atmospheres with absorption from H_2O , CO , CH_4 , NH_3 , CO_2 , or SO_2 . In particular, NIRSpec/PRISM’s sensitivity to CO_2 , which is out of the reach of NIRISS/SOSS, is essential because of its crucial role in dictating whether the conditions are met for surface liquid water on temperate rocky exoplanets (T. J. Fauchez et al. 2019). The fact that the atmospheric signal contained in the transmission spectrum seems flat, or muted, and without any haze-like slope enables us to place stringent constraints on the presence of an atmosphere on TRAPPIST-1 d.

All thick or cloud-free (cloud-top pressures deeper than 0.01 bar) hydrogen-rich ($\lesssim 100 \times$ solar metallicity) atmospheres are excluded by our observations at the $>3\sigma$ level. Since high-altitude aerosols are unlikely to form in hydrogen-rich temperate atmospheres (C. V. Morley et al. 2015; S. E. Moran et al. 2018), we rather explore two main possibilities: either TRAPPIST-1 d has an extremely thin, or cloudy, high mean molecular weight atmosphere, or it is airless.

For high mean molecular weight scenarios, we demonstrate that the precision reached by our observations can rule out compositions similar to solar system terrestrials: a cloud-free Titan composition, or even compositions analogous to Archean Earth, early Mars, or a cloud-free Venus are ruled out by our observations with greater than 95% confidence (2σ). Meanwhile, our spectra remain marginally consistent within the uncertainties (near the 2σ level) with an extremely thin Mars-like CO_2 atmosphere, a Venus-like scenario with high-altitude clouds, or the CO_2 signature of a modern Earth composition when accounting for Earth’s water clouds.

Our observations cannot yet completely exclude other potential atmosphere scenarios for TRAPPIST-1 d which were predicted in the literature: for instance, 3D GCM modeling demonstrates that a water-rich atmosphere may form high-altitude \sim millibar water clouds that reach the terminator, obscuring any features in transmission (T. J. Fauchez et al. 2019; M. Turbet et al. 2023). This issue is likely to impact atmosphere detectability in transmission for most HZ terrestrial planets (T. D. Komacek et al. 2020). As an alternative to transmission spectroscopy, reconnaissance eclipse photometry using JWST/MIRI’s 15 μm bandpass could confirm the ~ 200 ppm most optimistic airless scenario (i.e., no heat redistribution, zero-albedo surface) at 3σ confidence with seven eclipse

observations (R. Doyon 2024). Further modeling of the 3D impact of the condensation of water, as well as carbon-, nitrogen-, and sulfur-bearing species will be crucial for the interpretation of future transmission and emission studies of terrestrial exoplanets.

If TRAPPIST-1 d is airless, our evolution models predict that TRAPPIST-1 b, c, and d must have formed relatively dry, with an initial water inventory of less than about four Earth oceans. Because of its higher likelihood of atmospheric retention, atmospheric reconnaissance of TRAPPIST-1 d provides more leverage on the inner planets’ initial volatile inventories and the outer planets’ likelihood of atmosphere presence than TRAPPIST-1 b and c (vulnerable to complete atmosphere loss even if they initially accreted about 100 Earth oceans; J. Krissansen-Totton 2023). We find that even complete atmosphere loss for TRAPPIST-1 d would not preclude atmosphere presence for the outer HZ planets TRAPPIST-1 e, f, and g, as, contrary to the inner planets, they could retain water thanks to efficient shielding in the interior and in the condensed surface liquid water reservoir even if they initially accreted only a few Earth oceans of volatiles.

Acknowledgments

We thank the anonymous referee for comments that improved the quality of this manuscript. C.P.-G. also thanks D. Samra and J. Bean for helpful discussions. This work is based on observations with the NASA/ESA/CSA James Webb Space Telescope, obtained at the Space Telescope Science Institute (STScI) operated by AURA, Inc. This project was undertaken with the financial support of the Canadian Space Agency. All of the data presented in this paper were obtained from the Mikulski Archive for Space Telescopes (MAST) at the Space Telescope Science Institute. The data used in this paper can be found in MAST at the following DOI: [10.17909/2jzw-7m72](https://doi.org/10.17909/2jzw-7m72). C.P.-G. acknowledges support from the NSERC Vanier scholarship and the Trottier Family Foundation. C.P.-G. also acknowledges support from the E. Margaret Burbidge Prize Postdoctoral Fellowship from the Brinson Foundation. M.R. acknowledges funding from NSERC, Fonds de recherche du Québec—Nature et technologies (FRQNT), and iREx. R.A. acknowledges the SNSF support under the Post-Doc Mobility grant P500PT_222212 and the support of iREx. M.T. acknowledges support from the Tremplin 2022 program of the Faculty of Science and Engineering of Sorbonne University. M.T. acknowledges support from BELSPO BRAIN (B2/212/PI/PORTAL). D.B., K.M., and M.K. acknowledge support from InitiaSciences scholarships within the InitiaSciences research mentoring program. O.L. acknowledges financial support from the FRQNT. A.L’H. acknowledges support from the FRQNT under file #349961. L.D. is a Banting and Trottier Postdoctoral Fellow and acknowledges support from the Natural Sciences and Engineering Research Council (NSERC) and the Trottier Family Foundation. J.T. acknowledges funding support by the TESS Guest Investigator Program G06165. This project has been carried out within the framework of the National Centre of Competence in Research PlanetS supported by the Swiss National Science Foundation under grant 51NF40_205606. S.P. acknowledges the financial support of the SNSF. D.J. is supported by NRC Canada and by an NSERC Discovery Grant.

Software: Eureka!, jwst,²⁹ emcee³⁰ (D. Foreman-Mackey et al. 2013), corner³¹ (D. Foreman-Mackey 2016), batman³² (L. Kreidberg 2015), PHOENIX (T.-O. Husser et al. 2013), MSG³³ (R. Townsend & A. Lopez 2023), astropy³⁴ (Astropy Collaboration et al. 2013), numpy³⁵ (C. R. Harris et al. 2020), stctm³⁶ (C. Piaulet-Ghorayeb 2024, 2025, submitted), matplotlib³⁷ (J. D. Hunter 2007), scipy³⁸ (P. Virtanen et al. 2020), FastChem³⁹ (J. W. Stock et al. 2018).

Appendix A Stellar and Planetary Parameters

We present the key stellar and planetary parameters used in this work and inferred from our analysis in Table A1.

Table A1
Stellar and Planetary Parameters Adopted or Refined in This Work

| Parameter | Value | Source |
|--|------------------------------|--|
| TRAPPIST-1 (star) | | |
| Spectral type | M8.0V | J. Liebert & J. E. Gizis (2006), M. Gillon et al. (2017) |
| Effective temperature T_{eff} [K] | 2566 ± 26 | E. Agol et al. (2021) |
| Metallicity [Fe/H] | 0.052 ± 0.073 | F. Davoudi et al. (2024) |
| Age [Gyr] | 7.6 ± 2.2 | A. J. Burgasser & E. E. Mamajek (2017) |
| Stellar mass M_* [M_{\odot}] | 0.0898 ± 0.0023 | A. W. Mann et al. (2019) |
| Stellar radius R_* [R_{\odot}] | 0.1192 ± 0.0013 | E. Agol et al. (2021) |
| $\log_{10}(g_*)$ [cm s^{-2}] | $5.2396^{+0.0056}_{-0.0073}$ | E. Agol et al. (2021) |
| Stellar luminosity L_* [L_{\odot}] | 0.000553 ± 0.000019 | E. Agol et al. (2021) |
| Stellar rotation period [day] | 3.30 ± 0.14 | R. Luger et al. (2017) |
| TRAPPIST-1 d (planet) | | |
| Planetary radius R_p [R_{\oplus}] | $0.804^{+0.014}_{-0.012}$ | This paper |
| Planetary mass M_p [M_{\oplus}] | 0.388 ± 0.012 | E. Agol et al. (2021) |
| Orbital period P [day] | 4.049219 ± 0.000026 | E. Agol et al. (2021) |
| Scaled semimajor axis a/R_* | 40.15 ± 0.18 | This paper |
| Orbital separation a [au] | 0.0223 ± 0.0003 | Derived, this paper |
| Impact parameter b | 0.06 ± 0.05 | This paper |
| Inclination i [deg] | $89.909^{+0.068}_{-0.071}$ | Derived, this paper |
| Insolation S_p [S_{\oplus}] | $1.119^{+0.047}_{-0.044}$ | Derived, this paper |
| Equilibrium temperature | ... | ... |
| - $T_{\text{eq,A}} = 0.0$ [K] | 286 ± 3 | Derived, this paper |
| - $T_{\text{eq,A}} = 0.3$ [K] | 262 ± 3 | Derived, this paper |

Note. We combine the constraints on the planet’s radius and orbital parameters from both NIRSpec/PRISM visits to obtain the values quoted here. The derived values are obtained using Monte Carlo sampling. The quoted orbital parameters are from the E. Agol et al. (2021) correspond to the osculating Jacobi elements inferred from a transit timing analysis at the start time: $\text{BJD}_{\text{TDB}} - 2450000 = 7257.93115525$ days, and derived parameters are from a photo-dynamical analysis. Since the planet’s orbit evolves over time, the quoted values of the orbital parameters obtained from our fit to the two NIRSpec/PRISM transits can be considered to be a measurement of the planet’s orbital separation and inclination at the time of the observations.

²⁹ <https://github.com/spacetelescope/jwst>

³⁰ <https://github.com/dfm/emcee>

³¹ <https://github.com/dfm/corner.py>

³² <https://github.com/lkreidberg/batman>

³³ <https://github.com/rhdtownsend/msg>

³⁴ <https://www.astropy.org/>

³⁵ <https://github.com/numpy/numpy>

³⁶ <https://github.com/cpiaulet/stctm>

³⁷ <https://github.com/matplotlib/matplotlib>

³⁸ <https://github.com/scipy/scipy>

³⁹ <https://github.com/NewStrangeWorlds/FastChem>

Appendix B Light-curve Analysis Methodology

B.1. Broadband Light-curve Fitting

We construct our white light curves for each visit by summing all spectroscopic light curves, after discarding all wavelength channels where pixels are flagged as saturated in the median-in-time image. We find that this reduces the point-to-point scatter in the broadband light curve by $\approx 15\%$ to 18% in both visits. Raw white and spectroscopic light curves are shown in Figure B1. We then perform individual broadband light-curve fits to each light curve using the EXOTEP pipeline (B. Benneke et al. 2017, 2019a). We find that the light curves of both visits exhibit correlated noise on top of the transit signal. Although the light-curve variations do not track with

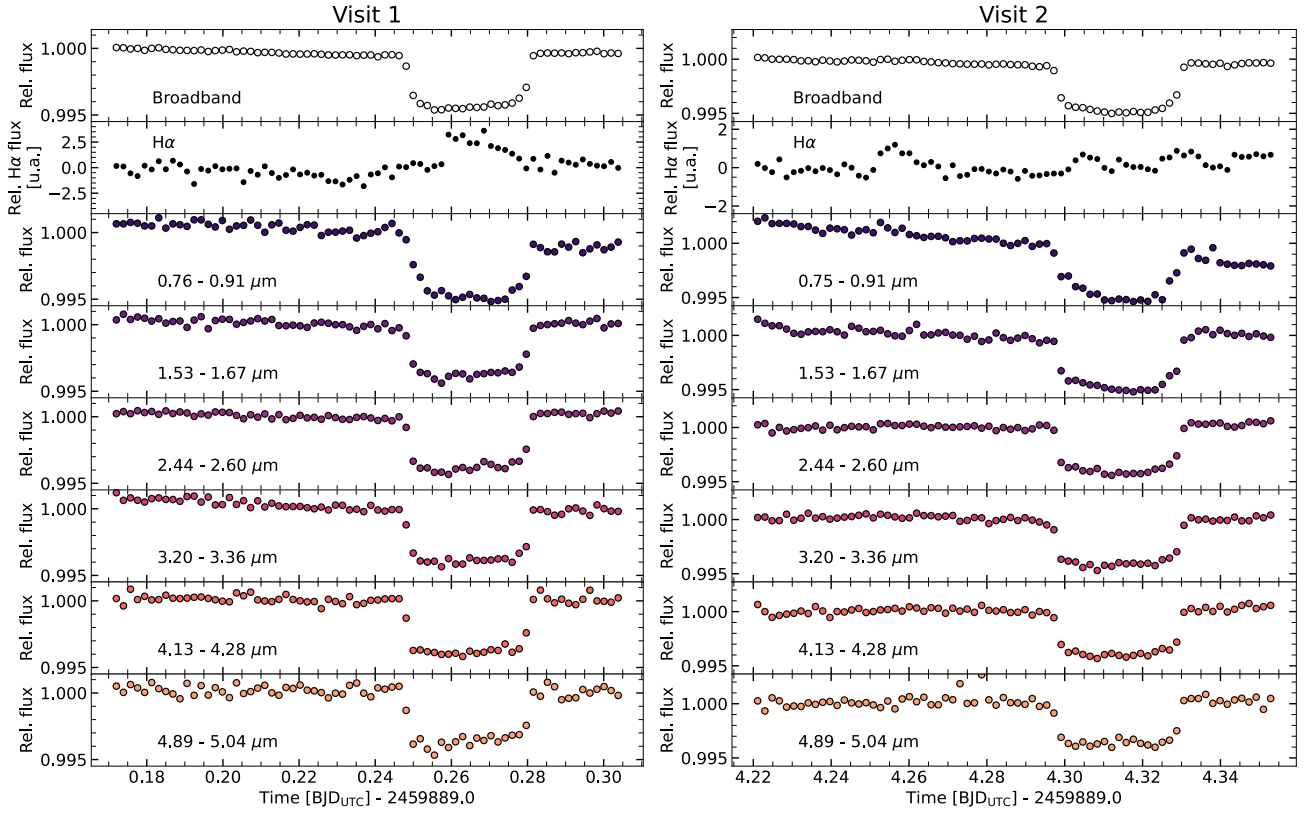


Figure B1. For visit 1 (left) and visit 2 (right), we show the raw broadband light curves (top panels), along with the $H\alpha$ flux (in arbitrary units, scaled by $\times 10^4$; second panels) and a few example raw spectroscopic light curves (bottom panels, with wavelength bins labeled in the bottom left corners). For clarity, we bin 100 data points together to obtain the curves shown in this figure. The spectroscopic light curves exhibit different slopes (generally more downward for shorter wavelengths) and wavelength-dependent features correlating with the small flares flagged in the $H\alpha$ time series.

Table B1
Fitted Planetary Parameters from the EXOTEP White Light-curve fits of Both JWST NIRSpec/PRISM Transits of TRAPPIST-1 d

| Fitted parameters | Visit 1 | Visit 2 |
|---|---|--|
| Midtransit time T_c [BJD _{TDB}] | 2459889.26447 \pm 0.00002 | 2459893.31397 \pm 0.00003 |
| Planet-to-star radius ratio R_p/R_* | 0.0614 \pm 0.0005 | 0.0624 ^{+0.0007} _{-0.0006} |
| Semimajor axis a/R_* | 40.16 ^{+0.08} _{-0.14} | 40.14 ^{+0.12} _{-0.22} |
| Impact parameter b | 0.06 ^{+0.04} _{-0.03} | 0.07 ^{+0.06} _{-0.05} |

Note. The planet-to-star radius ratio, scaled semimajor axis, and impact parameter are consistent within their uncertainties for both visits. The impact parameter was fitted with a prior of $\mathcal{N}(0.063, 0.063)$ obtained by E. Agol et al. (2021).

instrumental metrics such as the position or width of the trace in the spatial or dispersion direction, they correlate well with the recorded variations in the integrated flux around the $H\alpha$ line (Figure 1). As with previous studies of TRAPPIST-1 (O. Lim et al. 2023; W. S. Howard et al. 2023; M. Radica et al. 2025), we therefore attribute the variations we observe to stellar flares, although in our case the flares have very small amplitudes in terms of their imprint on the white light curve. They still impact our analysis given their unfortunate timing in and near the transit, suggesting that even small flare events can be problematic when extracting small-planet transmission spectra around late-type M dwarfs.

Direct detrending of the light curves against the $H\alpha$ flux time series itself was unsuccessful, as we observe flare-dependent time delays between the appearance of the flare in $H\alpha$ and the corresponding rise in flux observed in the white light curves. This is not surprising, given that flares have time

lags between the release of energy at different wavelengths (W. S. Howard et al. 2023). We therefore use a GP to model this astrophysical noise using a nonparametric model.

We test two GP kernels: the Matérn 3/2 kernel, and a simple harmonic oscillator (SHO) kernel with the Q parameter fixed to $1/\sqrt{2}$ (critically damped SHO term, as in, e.g., M. Radica et al. 2024). The Matérn kernel is a stationary kernel that generalizes the squared-exponential kernel with a parameter ν , which controls the smoothness of the resulting function. While $\nu = 1/2$ corresponds exactly to the squared-exponential kernel, $\nu = 3/2$ is the variant that produces once differentiable functions and is implemented in `celerite` (D. Foreman-Mackey et al. 2017). The critically damped SHO kernel, on the other hand, is more appropriate to model granulation noise than one-off flare events (L.-P. Coulombe et al. 2025), but we test it because of the small amplitude of the flares we model, and for comparison with other terrestrial M

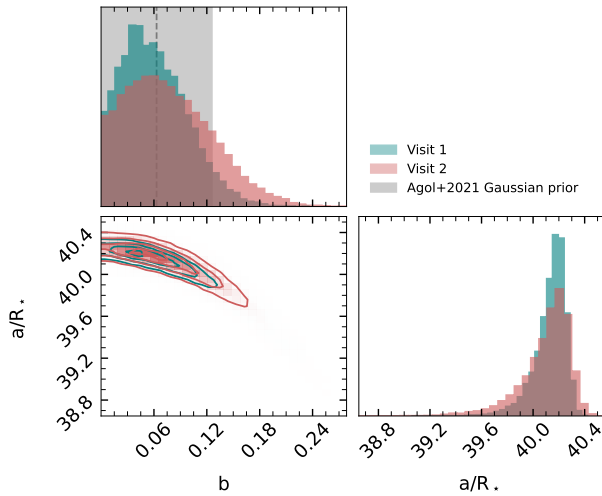


Figure B2. Joint (bottom left) and marginalized (top left, and bottom right) posterior distributions of the impact parameter b and the scaled semimajor axis a/R_* from the white light-curve fits to both visits (colored joint and marginalized distributions). The median and 1σ uncertainty region corresponding to the Gaussian prior on the impact parameter are indicated by the gray-shaded region.

dwarf planet light curves where it was previously applied (C. Cadieux et al. 2024; M. Radica et al. 2025).

Our full systematics model combines a linear slope with time, frequently observed in other JWST/NIRSpec data sets (e.g., Z. Rustamkulov et al. 2023), the GP parameters, and an extra jitter term added in quadrature to the diagonal of the covariance matrix. For each GP kernel, we vary the correlation amplitude parameter, as well as the correlation timescale in the fit.

In our astrophysical model, we fit the planet-to-star radius ratio, the time of transit center, the planet’s impact parameter, and its scaled semimajor axis (Table B1). After testing limb-darkening parameterizations where we fit a quadratic law with either the (q_1, q_2) triangular sampling basis (D. M. Kipping 2013) or the (u_1, u_2) uniform basis (see the following subsection for details), we adopt as fixed limb-darkening coefficients the best-fitting values obtained from the fit to the first NIRSpec/PRISM transit of TRAPPIST-1 g (B. Benneke et al. 2025, submitted), which was not affected by flares. As in O. Lim et al. (2023), we use a wide Gaussian prior on the impact parameter of $\mathcal{N}(0.063, 0.063)$ informed by previous photo-dynamical modeling of the TRAPPIST-1 planets (E. Agol et al. 2021); even with our good time sampling of the light curve, visit 2 in particular suffers from small flares near ingress and egress that could alter the inferred transit duration (Figure B2). We adopt wide uninformative priors on all the other parameters. Our final model combines the astrophysical (transit) component, and the systematics model described above. This full model provides a satisfactory fit to both light curves (Figure 1), with Gaussian-distributed residuals. We obtain consistent fit quality and inferred planetary parameters regardless of the GP kernel used, but we choose to adopt the results obtained with the Matérn 3/2 GP, since its covariance structure is more appropriate to model the one-off flare events we observed. A more detailed exploration of the impact of GP model assumptions is offered in the next subsection.

Table B2
NIRSpec/PRISM Spectrum of TRAPPIST-1 d for Visits 1 and 2

| Wavelength (μm) | Visit 1 | | | Visit 2 | | |
|---------------------------------|----------------|----------------------|----------------------|----------------|----------------------|----------------------|
| | Depth (ppm) | +1 σ (ppm) | −1 σ (ppm) | Depth (ppm) | +1 σ (ppm) | −1 σ (ppm) |
| 0.60–0.79 | 3407.6 | 396.4 | 345.7 | 2919.2 | 394.9 | 438.2 |
| 0.79–0.86 | 3468.6 | 158.5 | 155.0 | 3275.7 | 108.2 | 95.6 |
| 0.86–0.94 | 3544.3 | 96.9 | 96.0 | 3465.8 | 95.4 | 93.3 |
| 0.94–1.02 | 3554.6 | 134.6 | 153.1 | 3608.7 | 95.7 | 101.7 |
| 1.02–1.11 | 3742.3 | 64.3 | 63.1 | 3788.6 | 106.1 | 102.3 |
| 1.11–1.22 | 3618.3 | 73.4 | 76.1 | 3697.0 | 107.6 | 93.1 |
| 1.22–1.33 | 3721.4 | 86.0 | 85.1 | 3791.7 | 139.9 | 158.9 |
| 1.33–1.47 | 3660.3 | 88.7 | 84.4 | 3963.5 | 125.8 | 166.4 |
| 1.47–1.63 | 3494.0 | 112.5 | 104.5 | 3784.4 | 89.6 | 88.6 |
| 1.63–1.75 | 3415.0 | 117.2 | 137.0 | 3920.4 | 122.5 | 112.7 |
| 1.75–1.89 | 3524.4 | 118.6 | 117.6 | 3933.1 | 141.9 | 124.7 |
| 1.89–2.04 | 3519.3 | 113.4 | 105.6 | 3819.2 | 123.5 | 103.4 |
| 2.04–2.20 | 3668.8 | 130.3 | 118.0 | 3958.6 | 134.6 | 131.7 |
| 2.20–2.38 | 3566.2 | 84.6 | 79.0 | 3799.8 | 120.1 | 119.8 |
| 2.38–2.57 | 3605.2 | 77.7 | 75.7 | 3978.6 | 108.3 | 111.6 |
| 2.57–2.78 | 3734.5 | 135.9 | 125.1 | 3874.1 | 80.7 | 86.5 |
| 2.78–3.00 | 3742.5 | 143.3 | 123.6 | 3843.1 | 129.6 | 113.4 |
| 3.00–3.24 | 3667.5 | 128.1 | 116.3 | 3883.0 | 99.1 | 99.2 |
| 3.24–3.50 | 3529.4 | 92.7 | 81.3 | 3731.5 | 124.8 | 104.9 |
| 3.50–3.79 | 3606.2 | 99.2 | 94.2 | 3918.2 | 103.2 | 108.1 |
| 3.79–4.09 | 3564.3 | 99.5 | 91.9 | 3960.4 | 101.6 | 95.2 |
| 4.09–4.41 | 3682.4 | 107.1 | 103.8 | 3845.0 | 82.7 | 109.2 |
| 4.41–4.76 | 3750.4 | 68.2 | 68.0 | 3674.6 | 113.3 | 89.8 |
| 4.76–5.20 | 3689.0 | 134.6 | 110.9 | 3562.4 | 75.6 | 75.2 |

Note. These spectra are obtained directly from the transit light curves, without correction for the impact of stellar contamination. A machine-readable version of this table will be made available in the electronic journal.

(This table is available in machine-readable form in the [online article](#).)

B.2. Spectroscopic Light-curve Fitting

We obtain a transmission spectrum of TRAPPIST-1 d for each visit by fitting the spectroscopic light curves using EXOTEP, and explore the impact of our choice of limb-darkening and GP kernel prescriptions on the final spectrum. We find that flares impact the determination of the limb-darkening parameters in the second visit, and that the correlated noise we observe has chromaticity that needs to be accounted for.

In all the fits described below, we follow a similar procedure to the white light-curve analysis and jointly fit the astrophysical transit model described by the planet-to-star radius ratio R_p/R_* , and a systematics model including parameters that capture instrumental and astrophysical noise, to each spectroscopic light curve. We fix the orbital parameters (b and a/R_*), and the midtransit time T_c to the best-fit parameters from the white light-curve fit.

Our first step consists in obtaining a first-pass spectrum at the full instrumental resolution, with one light curve extracted for each column along the dispersion direction. The goal of this step is to identify any outlier columns that should be discarded for the final analysis, especially in the region of the detector where ramps were affected by saturation. For this spectrum, we adopt EXOTIC-LD (T.-O. Husser et al. 2013; D. Grant & H. R. Wakeford 2024) 1D limb-darkening coefficients, as our light curves do not have the S/N required to fit them. We choose this approach despite the fact that stellar model fidelity can be an issue for late M dwarfs since

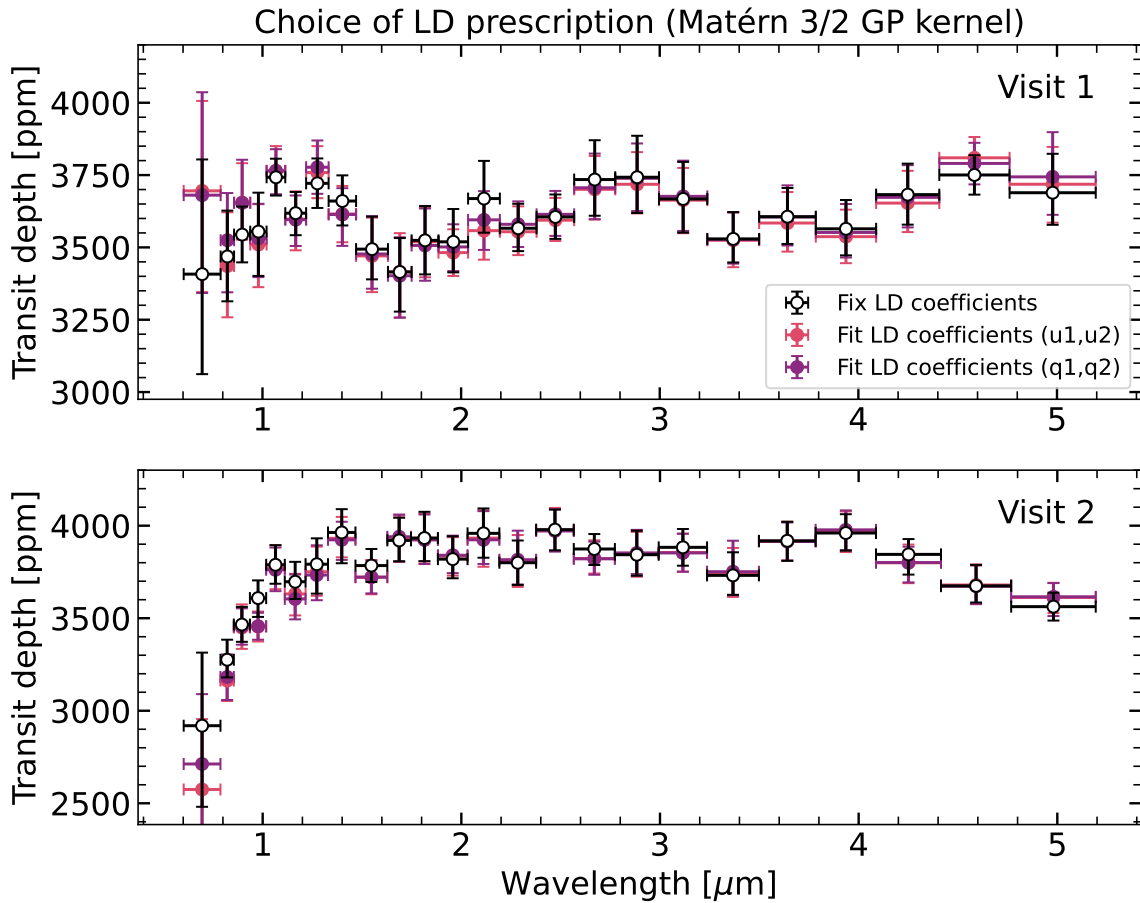


Figure B3. Assessment of the impact of either fixing the limb-darkening parameters to the best-fit values from the first NIRSpec/PRISM transit of TRAPPIST-1 g (B. Benneke et al. 2025, submitted; shown in black) or fitting them in the broadband and spectroscopic light-curve fits using the $[u_1, u_2]$ basis (pink) or the $[q_1, q_2]$ parameterization (magenta). The top (bottom) panel shows the results for the first (second) visit. Our spectra are largely insensitive to the choice of limb-darkening parameterization, except at short wavelengths in both visits (more pronounced for visit 2). The presence of small stellar flares during the first transit and near the ingress and egress of the second transit likely alter the determination of the limb-darkening parameters from the light curves at short wavelengths.

our aim is to identify columns that are outliers relative to their neighbors rather than to extract a valid spectrum for further analyses. We fit a simple transit model with a slope as the systematics model. We find that the following outlier columns should be discarded, as they were $>4\sigma$ outliers to the neighboring pixels: for visit 1, the column centered at $1.4115 \mu\text{m}$, for visit 2, the column centered at $1.0663 \mu\text{m}$, and for both visits, the three columns spanning 1.5716 to $1.6117 \mu\text{m}$. To obtain the final spectra, we therefore construct the binned light curves by ignoring these four columns in each visit.

We perform sensitivity analyses to evaluate the impact of the choice of the limb-darkening parameters and of the correlated noise model on our transmission spectrum. We use as a benchmark for the limb-darkening coefficients the values that were fitted to the first NIRSpec/PRISM transit of TRAPPIST-1 g (B. Benneke et al. 2025, submitted), as this observation was not affected by small or large flares that could have a wavelength-dependent impact on the light curves and the determination of the star’s intensity profile (and is largely free of unocculted stellar surface heterogeneity signatures). In order to leverage these limb-darkening coefficients, we bin the light curves following the wavelength bins of B. Benneke et al. (2025, submitted) prior to the spectroscopic fit.

B.2.1. Limb-darkening Prescription

We evaluate the impact of fitting, or fixing, the limb-darkening coefficients to independently determined values on the final transmission spectrum. For this series of tests, the systematics model consists in a slope with time, a Matérn 3/2 GP where the GP amplitude and timescale are fitted for each wavelength bin (an assumption that is explored further in the next subsection), and an additional white noise scatter term. The astrophysical transit model describing each of the spectroscopic light curves has either one or three fitted parameters: R_p/R_* , and the two coefficients of a quadratic limb-darkening law if they are fitted.

We perform three tests: (1) holding the limb-darkening parameters fixed to the best-fit values fitted for TRAPPIST-1 in each of the same wavelength bins from B. Benneke et al. (2025, submitted); (2) fitting the quadratic limb-darkening parameters q_1 and q_2 (D. M. Kipping 2013) in each bin; and (3) fitting instead u_1 and u_2 in order to alleviate a potential bias from the (q_1, q_2) prior on the transmission spectrum (L.-P. Coulombe et al. 2024).

We find excellent agreement between the three versions of the transmission spectrum for each visit (Figure B3), which suggests that at least at wavelengths longer than $\sim 1.5 \mu\text{m}$, we can reliably constrain the limb-darkening parameters from our

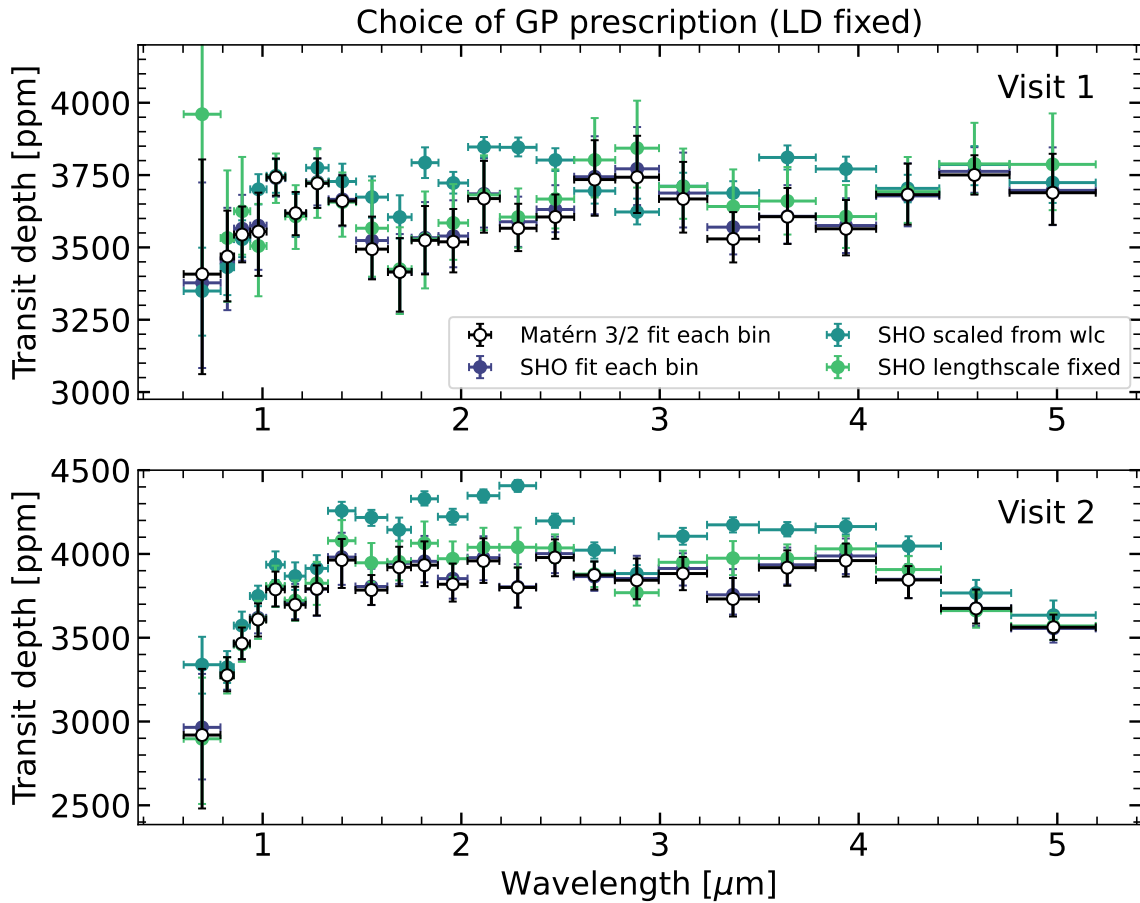


Figure B4. Assessment of the impact of the GP parameterization adopted for the spectroscopic fits, for visit 1 (top) and visit 2 (bottom), in the case where the limb-darkening parameters are kept fixed to prescribed values (see Figure B3). We tested two different GP kernels: the Matérn 3/2 kernel (black), and the SHO kernel (shades of blue). When the GP amplitude and length scale are fitted in each wavelength bin (black and dark blue points), the final spectrum is virtually identical. Assuming that the GP kernel length scale that corresponds to the best-fitting white light-curve model is also shared across all wavelength bins yields a consistent spectrum to a case where it is fitted, but the light-curve fit quality from 1.5 to 4 μm is improved by fitting independent values (light green). Finally, the best-fitting GP component from the white light-curve fit does not match well the spectroscopic light curves over most of the wavelength range (dark green), and leads to underestimated error bars because of the assumed correlated noise structure.

spectrum. However, we note a discrepancy at the shorter-wavelength end, $\lambda \lesssim 1.5 \mu\text{m}$, resulting in systematically lower inferred transit depths in the second visit when the limb-darkening parameters are fitted, compared to when they are fixed. Such short-wavelength differences may be driven by flares that affect the spectrum more strongly at shorter wavelengths (W. S. Howard et al. 2023) and are observed coinciding with the timing of ingress and egress for the second TRAPPIST-1 d visit. Since this difference with the fixed limb-darkening version is not observed for the first visit, while a real offset would be systematic for a shared host star, we elect to use the spectrum obtained with the limb-darkening coefficients fitted to the transit of TRAPPIST-1 g.

B.2.2. Gaussian Process Parameterization

The treatment of correlated noise at the spectroscopic light-curve fit stage relies on our assumptions about the chromaticity and functional approximation of the flare signature. We test several scenarios.

1. *Achromatic signature.* Scaling the mean GP model extracted from the best fit to the white light curve for each spectroscopic light curve, maintaining a mean of zero

for the systematics model, but fitting a multiplicative parameter that allows us to vary the amplitude of the variations relative to zero residuals.

2. *Achromatic correlation timescale.* Fitting the amplitude of the GP to each spectroscopic light curve individually, but keeping the GP timescale fixed to the best-fit value from the white light-curve fit. In this scenario, contrary to the first one, the correlation structure of the residuals is inferred from each spectroscopic light curve (using the positions of neighboring points) rather than by the white light-curve residuals.
3. *Chromatic timescale and amplitude.* Fitting both the GP amplitude and timescale in each spectroscopic bin. For this scenario, we tested the impact of switching the GP kernel between the Matérn 3/2 and SHO kernel.

For each of these tests, we hold the limb-darkening parameters fixed to the B. Benneke et al. (2025, submitted) values.

We find that, while approaches (2) and (3) provide consistent spectra, method (1) yields a very different spectrum (Figure B4). Therefore, we infer that the timescale over which the flare affects the light curves is largely wavelength independent, but that the exact timing and shape of the flare signal is chromatic (Figure B1). The choice of GP kernel does not affect the spectrum, and we choose to use the Matérn

3/2 GP result in the rest of the analysis as it is inherently tailored to match the type of nonperiodic flare signal we are correcting. Finally, we find that when fitting the GP timescale in each wavelength bin, it can be satisfactorily constrained by each individual light curve. Since this approach better captures the uncertainty on the GP timescale rather than assuming one value extracted from the white light curve, we move forward with the spectrum obtained with method (3) (see fitted spectroscopic light curves in Figure 1). The transmission spectra we obtain are provided in Table B2.

Appendix C

Out-of-transit and In-transit Constraints on Stellar Heterogeneities

We constrain the coverage of bright and dark heterogeneities on the visible stellar hemisphere from two independent methods: retrievals on each visit’s out-of-transit stellar spectra, and on the planet’s transmission spectra affected by unocculted stellar surface heterogeneities.

Throughout this study, we use the words “facula(e)” (“spot (s)”) to refer to regions of the stellar surface that are well represented by the 1D spectrum of a hotter (cooler) star than the photosphere, i.e., brighter or dimmer at all wavelengths (following, e.g., B. V. Rackham et al. 2018; A. R. Iyer & M. R. Line 2019). However, we recognize that this terminology is limited since the nature of stellar surface heterogeneities on late-type M dwarf stars can be drastically different than their F, G, or even K dwarf counterparts. For example, modeling work (see, e.g., B. Beeck et al. 2015; C. M. Norris et al. 2023) has suggested that magnetic activity could result in the formation of dark faculae with bright hot walls on the surfaces of M dwarfs in between granules. Such 3D modeling of facula contrast spectra also revealed that they can be both brighter and dimmer than the quiet photosphere depending on the wavelength range considered. This means that our simplified models leveraging 1D grids may lead to skewed inferences on the coverage fractions of different heterogeneities at the stellar surface, and we recognize that the interpretability of our results would benefit from further model fidelity. However, without any large grids of late M heterogeneity contrast spectra available, as required to properly model these effects, we default to the current state-of-the-art method for stellar surface heterogeneity considerations in transit spectroscopy as described in Appendices C.2 and C.3.

C.1. Flux-calibrated Out-of-transit Stellar Spectrum

For each visit, we perform a modified version of the data reduction in order to obtain a time series of flux-calibrated stellar spectra. The steps are identical to the standard data reduction described in Section 2.2, except for a few differences in Stages 2 and 3. Stage 2 is performed using *Eureka!*, except that the flat-field and absolute photometric calibration steps (*FlatFieldStep* and *PhotomStep* in the *jwst* pipeline, respectively) are not skipped. We also modify the Stage 3 *Eureka!* code where the optimal spectral extraction is performed in order to mask rows that are more than 8 pixels away from the trace, following S. E. Moran et al. (2023).

We then select integrations from which to construct the out-of-transit spectra. Since our analysis aims at determining the combination of stellar photosphere and surface heterogeneity components that best describe the “quiet” photosphere, we discard not only in-transit integrations, but also out-of-transit

integrations which coincide with likely flare events from the $H\alpha$ time series and the shape of the fitted GP component (Figure 1). Specifically, for visit 1, we only consider integrations prior to $\text{BJD}_{\text{UTC}} = 2459889.245$, and for visit 2, we discard integrations in $\text{BJD}_{\text{UTC}} \in [2459893.245, 2459893.27]$ and $\text{BJD}_{\text{UTC}} > 2459893.295$. We then compute the median in time of all the spectra from the remaining integrations to create the out-of-transit spectrum for each visit, and set as the error bar on each point the standard deviation in that wavelength band over all the integrations considered. Finally, when performing the out-of-transit retrieval on the flux-calibrated spectra, we also crop all wavelengths greater than $5.5 \mu\text{m}$, which is the long-wavelength edge of the PHOENIX models (T.-O. Husser et al. 2013) that we use to construct the model spectra (see Section 4).

C.2. Out-of-transit Stellar Spectrum Modeling

exotune models the out-of-transit stellar spectrum as a linear combination of one, two, or three spectral components, each defined by its effective temperature T_{comp} and $\log g_{\text{comp}}$. The one-component model assumes that the entire spectrum is described by the quiet photosphere, while the two- or three-component models simulate the emergent spectrum as the linear combination of two or three stellar spectra, with one corresponding to the photosphere, and the other(s) describing a cooler “spot” component (covering a fraction f_{spot} of the stellar surface) and/or a hotter “facula” component (covering a fraction f_{fac} of the visible hemisphere on the stellar surface). Our model out-of-transit stellar flux F_{oot} is therefore constructed as

$$F_{\text{oot}} = (1 - f_{\text{spot}} - f_{\text{fac}}) \times F_{\text{phot}} + f_{\text{spot}} \times F_{\text{spot}} + f_{\text{fac}} \times F_{\text{fac}}, \quad (\text{C1})$$

where F_{phot} (T_{phot} , $\log g_{\text{phot}}$), F_{spot} (T_{spot} , $\log g_{\text{spot}}$), and F_{fac} (T_{fac} , $\log g_{\text{fac}}$) are stellar surface fluxes computed at $R = 10,000$ from the PHOENIX library (T.-O. Husser et al. 2013) obtained with the *MSG* module (R. Townsend & A. Lopez 2023). This forward model is implemented in *exotune* within a retrieval framework by precomputing a grid of stellar models in increments of 10 K in T_{eff} and 0.1 in $\log g$ and choosing at each iteration and for each component the closest model in the grid. Once the high-resolution models are combined to obtain F_{oot} , we integrate the flux within each wavelength bin of the NIRSpec/PRISM spectrum to compute the Gaussian likelihood within a Bayesian framework. We adopt uniform priors on f_{spot} and f_{fac} requiring $f_{\text{spot}} + f_{\text{fac}} < 1$, on $T_{\text{phot}} \in [2300, 3700]$ K (2300 K being the lowest temperature in the PHOENIX model grids), and on the temperature contrasts of the hot and cool heterogeneity components, $|\Delta T_{\text{het}}| \in [50, 1000]$ K ($\Delta T_{\text{fac}} > 0$, $\Delta T_{\text{spot}} < 0$, and $T_{\text{het}} = T_{\text{phot}} + \Delta T_{\text{het}}$). We also set uniform priors on the photosphere $\log g \in [2.5, 5.5]$ (since PHOENIX models with a value of $\log g \sim 3.5$, much lower than the literature value, were found to provide a better match to the NIRISS/SOSS transmission spectrum of TRAPPIST-1 from a forward model exploration; O. Lim et al. 2023; Figure C1). We similarly adopt a uniform prior on the difference $\Delta \log g_{\text{het}} < 0$ between the heterogeneity and the photosphere components; such a difference could be induced by magnetic activity (see, e.g., M. Fournier-Tondreau et al. 2023). Additionally, we fit a factor to scale F_{oot} to the flux of TRAPPIST-1 received at Earth,

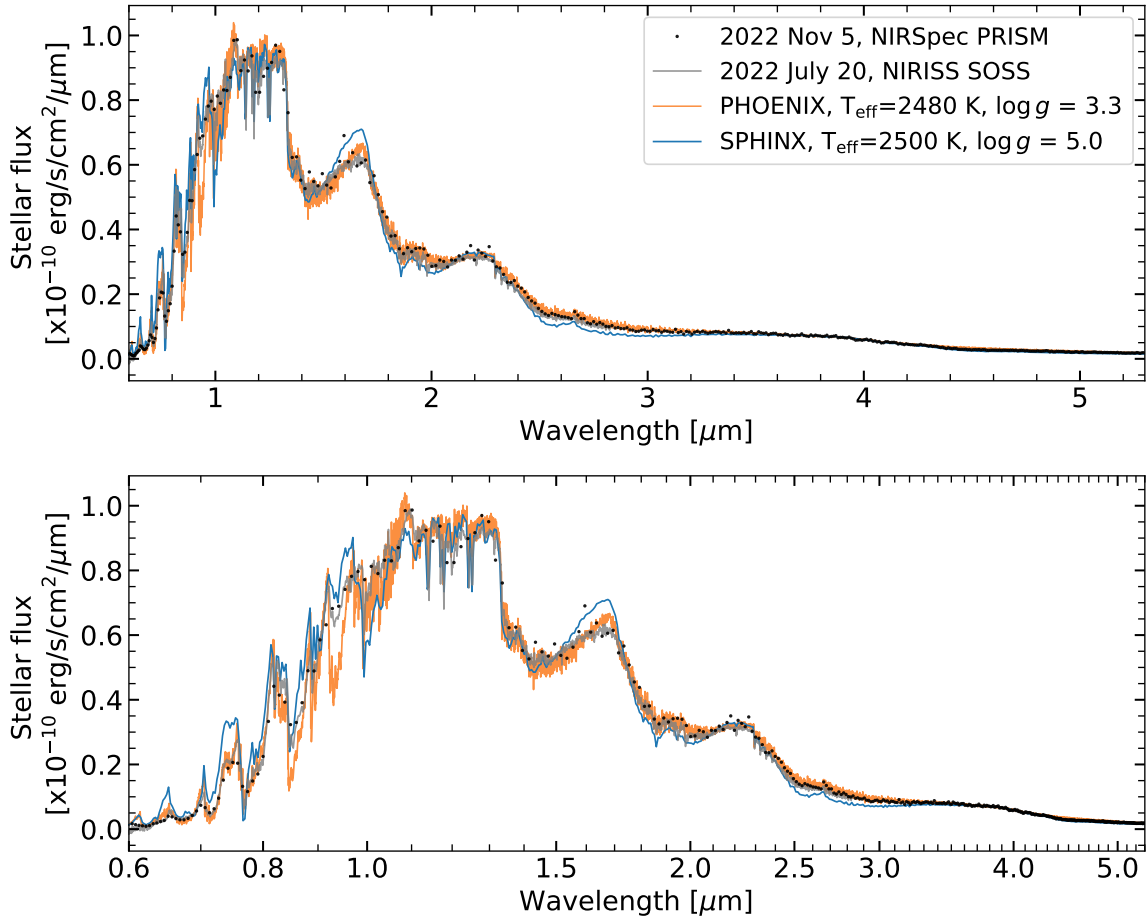


Figure C1. Comparison of the NIRSpec/PRISM spectrum of TRAPPIST-1 (black points) with NIRISS/SOSS observations (gray line; O. Lim et al. 2023) and stellar models. We show the PHOENIX model in our grid with the T_{eff} and $\log g$ combination that provides the closest match to our NIRSpec/PRISM spectrum in orange (convolved to $R = 2000$), and a SPHINX model with $T_{\text{eff}} = 2500$ K and $\log g = 5$ (A. R. Iyer et al. 2023a, 2023b). Both PHOENIX models and SPHINX models struggle to reproduce the short-wavelength part of the spectrum, and the SPHINX model additionally does not capture well the relative depths of the broadband spectral features observed past $1.5 \mu\text{m}$.

Table C1

Results from the Three-component Retrievals Performed with `stctm` on the (In-transit) Transmission Spectra, and on the Out-of-transit Stellar Spectra

| Parameter | Visit 1 | | Visit 2 | |
|------------------------------|----------------------------|----------------------------|----------------------------|----------------------------|
| | In-transit (TLS) | Out-of-transit | In-transit (TLS) | Out-of-transit |
| T_{phot} [K] | $2604.0^{+67.0}_{-64.0}$ | $2630.0^{+93.0}_{-69.0}$ | $2557.0^{+60.0}_{-54.0}$ | $2639.0^{+111.0}_{-72.0}$ |
| $\log g_{\text{phot}}$ | $3.4^{+0.9}_{-0.4}$ | $4.0^{+0.6}_{-0.4}$ | $4.8^{+0.3}_{-0.5}$ | $4.0^{+0.7}_{-0.5}$ |
| T_{spot} [K] | $2388.0^{+90.0}_{-60.0}$ | $2418.0^{+62.0}_{-73.0}$ | $2357.0^{+58.0}_{-40.0}$ | $2415.0^{+59.0}_{-67.0}$ |
| f_{spot} | $0.18^{+0.25}_{-0.14}$ | $0.64^{+0.17}_{-0.32}$ | $0.40^{+0.18}_{-0.18}$ | $0.68^{+0.15}_{-0.32}$ |
| T_{fac} [K] | $3281.0^{+163.0}_{-248.0}$ | $3127.0^{+268.0}_{-272.0}$ | $2957.0^{+215.0}_{-150.0}$ | $3077.0^{+273.0}_{-249.0}$ |
| f_{fac} | $0.007-0.076$ | <0.10 | $0.058-0.274$ | <0.089 |
| $\Delta \log g_{\text{het}}$ | >-1.3 | >-2.1 | >-0.6 | >-2.2 |

Note. We report 1σ confidence intervals for each parameter, or 2σ upper or lower limits where applicable. We use the posterior samples of ΔT_{spot} , ΔT_{fac} , and T_{phot} to obtain the values quoted for T_{spot} and T_{fac} .

and an error inflation factor $f_{\text{errInfl}} \in [0, 50]$ that multiplies uniformly the estimated errors on the measured flux to account for a potentially imperfect model fidelity.

We explore the parameter space using the Markov Chain Monte Carlo sampler `emcee` (D. Foreman-Mackey et al. 2013), within a fully parallelized framework. We use 20 times as many walkers as there are fitted parameters. Each chain is

run for 100,000 steps, 60% of which are ignored (burn in) to obtain the final posterior distributions of the heterogeneity parameters. We ensure that the chains run for more than 50 times the largest autocorrelation timescale inferred across all parameters (<840 steps across all our runs). Motivated by the downward slopes in the transmission spectra at the blue end of the NIRSpec/PRISM spectrum (Figure 2), we perform two-

(photosphere and faculae) and three-component (photosphere, spots, and faculae) retrievals on the out-of-transit stellar spectrum for each visit. We record the Bayesian information criterion (BIC) value for each retrieval for model comparison purposes.

C.3. Transit Light Source Effect Modeling

We further use `stctm` (C. Piaulet-Ghorayeb 2024, 2025, submitted) to model the transmission spectrum of TRAPPIST-1 d, assuming that the planet is either a bare rock or that its transmission spectrum is otherwise featureless. We model the TLS effect assuming two populations of heterogeneities referred to as spots and faculae, and parameterized as previously described in C. Piaulet-Ghorayeb et al. (2024) and in the previous section. We also perform retrievals where only one population of heterogeneities (faculae) is considered, in order to assess the impact of degeneracies between spot and facula heterogeneity properties on the final posterior distributions we obtain.

In our retrievals, we fit the temperatures of the heterogeneity components following the same parameterization as in the `exotune` fit to the out-of-transit stellar spectrum. We similarly use a precomputed fine grid of stellar models to compute the forward models for each model evaluation.

The parameter space is sampled using `emcee` (D. Foreman-Mackey et al. 2013), using 20 times as many walkers as there are fitted parameters, and running the chains for 100,000 steps, with 60% discarded as burn in, confirming that each chain ran for more than 50 times the maximum estimated autocorrelation timescale (ranging from 500 to 1700 steps depending on the run). For the heterogeneity fractions, we test both uniform and log-uniform priors, always ensuring that $f_{\text{spot}} + f_{\text{fac}} < 1$ in order to assess whether the data could provide a lower bound on f_{fac} given the strong downward slope observed in the visit 2 spectrum (Figure 2). The model spectra and stellar heterogeneity parameters are consistent regardless of the prior we adopt, and we report the results from the fit performed with a log-uniform prior on the facula covering fraction for the two- and three-component retrievals (since we recover a lower limit for visit 2), and with a linear prior on the spot covering fraction (which is unconstrained regardless of the prior we adopt) for the three-component fit. We adopt uniform priors on the heterogeneity temperature contrast ΔT_{het} , with $|\Delta T_{\text{het}}| > 100$ K to ensure that the heterogeneity spectra do not default to the photosphere spectrum. Contrary to the out-of-transit stellar spectrum fit, the TLS retrievals are not sensitive to the photosphere temperature, but only to the ratios between the photosphere and heterogeneity component spectra. Therefore, we impose a Gaussian prior on T_{phot} using the median effective temperature value reported by E. Agol et al. (2021) and a wider standard deviation of 70 K, following the procedure used for the `stctm` fit to the transmission spectra of TRAPPIST-1 b (O. Lim et al. 2023) and c (M. Radica et al. 2025). We implement a new functionality in `stctm` that allows us to also fit the $\log g$ of the photosphere (same prior as the `exotune` retrieval) and the $\Delta \log g$ contrast between the surface gravity used to describe the heterogeneity and photosphere components, in a way analogous to `exotune`. Finally, we fit the wavelength-independent transit depth of the planet itself, $D_{\text{scale, TLS}}$, with a wide uniform prior. We record the BIC value in each case in order to evaluate to what extent higher-complexity models are favored by the data.

C.4. Constraints from the Transit Light Source Effect

Both spectra exhibit telltale signs of stellar contamination (Figure 2), in particular the downward slope shortward of about $1 \mu\text{m}$, which can only be reproduced by the dilution effect of unocculted bright inhomogeneities on the stellar surface. The good match to the spectra with the `stctm` retrievals support this, as we can reproduce most of the features without the need for a planetary atmosphere (Figure 2). The $4\text{--}5 \mu\text{m}$ slope in the visit 2 transmission spectrum could not be accurately matched by our models, and would also be surprising from an atmospheric standpoint. We do not find motivation in the spectroscopic light curves to discard the two redmost points. Still, only one point cannot be reproduced within 1σ by the `stctm` model, and we confirm that running the SCARLET atmosphere retrievals with or without that point does not substantially impact our inferences.

Overall, the visit 2 spectrum is the most affected by stellar contamination, with the measured transit depth at $0.7 \mu\text{m}$ being about 1000 ppm lower than that measured at $1.3 \mu\text{m}$. Beyond the short-wavelength slope, the median transit depth from 1.2 to $5 \mu\text{m}$ is also higher in the second visit compared to the first while we infer consistent orbital parameters across visits (Figure B2), supporting the interpretation of visit-to-visit variations in stellar contamination (Figure 2).

The short-wavelength slopes in the spectra favor the inclusion of faculae rather than assuming only a single-temperature photosphere (1.8σ for the visit 1 TLS fit, 2.8σ for the visit 2 retrieval), but the contribution of spots to the fit, especially for visit 2 where a large f_{spot} is favored, is nontrivial in three-component fits. We find that the ΔBIC disfavors adding a spot component for the first visit (the additional parameters are not warranted by the marginal improvement in fit quality). Meanwhile, the addition of spots to the visit 2 retrieval is only weakly supported with a ΔBIC of 2.3. Comparing the best-fit model from the three-component fits to a two-component (photosphere and faculae) alternative sheds some light on these results (Figure 2, top panel). Generally, for both visits, the two- and three-component models are identical shortward of about $1.3 \mu\text{m}$ (i.e., the downward slope is purely described by the facula component). However, the transit depth variations longward of $1.3 \mu\text{m}$ are better matched with the addition of the cool heterogeneity. For visit 1, spots provide a better match to the (tentative, $<1\sigma$) 1.5 to $2 \mu\text{m}$ dip and to the 2.5 to $3 \mu\text{m}$ bump in the spectrum, while the model accounting only for faculae produces a flat spectrum in these regions. Conversely, for visit 2, the facula-only model predicts, given the large f_{fac} , that one would observe TLS-imprinted “inverse” water features peaking near 1.4 , 1.8 , and $2.8 \mu\text{m}$, while the observed spectrum is much flatter, motivating the addition of the spot component.

C.5. Characteristics of the Out-of-transit Stellar Spectrum

The `exotune` out-of-transit stellar spectral retrievals for each individual visit yield consistent conclusions. Both visits can be explained by an ~ 2500 K photosphere, a minor facula component with 2σ upper limits on f_{fac} of about 10%, and a potential spot component. Since two-component (photosphere and faculae) retrievals provided a similarly good match to the data, while inferring lower photospheric temperatures closer to 2450 K, we hypothesize that in the case of the three-component fit, the preferred large spot covering fraction

essentially buffers for the fact that the bulk of the spectrum could be explained by a cool, 2450–2500 K component. For either visit, the ΔBIC does not significantly favor the inclusion of the spot component in addition to the faculae.

These conclusions hold whether or not we include in the *exotune* fit the columns affected by partial saturation. Over these wavelengths, our stellar spectrum is also consistent with the one reported with NIRISS/SOSS (O. Lim et al. 2023), which was not affected by saturation (see Figure C1).

C.6. Potential Caveats of the Stellar Surface Modeling

The reliability of stellar surface inferences for TRAPPIST-1 from stellar models using the PHOENIX model grid deserves further scrutiny. First, stellar model fidelity for late M dwarfs is imperfect, and may call for data-driven approaches given important mismatches between observations and models at the level of stellar SEDs, as well as missing lines in the reference line lists (B. V. Rackham et al. 2023; TRAPPIST-1 JWST Community Initiative et al. 2024; F. Davoudi et al. 2024; F. Jahandar et al. 2025).

We find, consistent with previous work (O. Lim et al. 2023), that we need to use the stellar photosphere $\log g$ as a fudge factor to accurately match the stellar spectrum with PHOENIX models (Figure C1). Both TLS and stellar spectral retrievals favor low $\log g \sim 3.5\text{--}4.0$ (Table C1) that are inconsistent with the literature inferences for TRAPPIST-1, but allow us to better reproduce the observed spectral features. We note that, although low $\log g$ values are favored by our analysis, the low resolution of NIRSpec/PRISM relative to NIRISS/SOSS reduces our sensitivity to the two main features near 1.05 and 1.6 μm that drive the inference on $\log g$ (Figure C1). Future studies aiming at characterizing the star should ideally use both NIRISS and NIRSpec in conjunction, in order to achieve high enough resolving power at the short wavelengths while capturing the long-wavelength Rayleigh–Jeans tail of the spectrum.

As an alternative to the PHOENIX models, we considered the SPHINX (A. R. Iyer et al. 2023a, 2023b) model grid. We compare our spectrum to a nominal SPHINX model for $T_{\text{eff}} = 2500$ K, $\log g = 5.0$, and solar $[\text{Fe}/\text{H}]$ and C/O ratios, and find that although some spectral regions near 1.1–1.3 μm are slightly better fit by this model, others at shorter wavelengths show a similar mismatch to the data as the PHOENIX models do, and further fail to reproduce the relative strengths of the broadband spectral features observed past 1.3 μm . We do not perform retrievals using SPHINX models, since their $R \sim 250$ resolving power is too low to be further leveraged by joint TLS–atmosphere retrievals, which reduces their interest in terms of contextualizing findings about the extent to which stellar contamination affects the transmission spectrum.

We also recognize as a potential caveat the fact that the stellar effective temperature, near 2600 K, is only 300 K away from the lowest temperature covered by the PHOENIX model grids, at 2300 K. More detailed models tailored to the conditions on low-mass M dwarfs could therefore help tease out more stellar surface information content from future observations.

Appendix D

Details of the Atmosphere Modeling Methodology

D.1. Sequential Retrievals of Atmospheric Properties

We use the SCARLET (B. Benneke & S. Seager 2012; B. Benneke 2013, 2015; B. Benneke et al. 2019a, 2019b; S. Pelletier et al. 2021; C. Piaulet et al. 2023) 1D atmospheric forward modeling and retrieval framework to interpret the observed transmission spectrum of TRAPPIST-1 d. With SCARLET, we fit the combined (visit 1 + visit 2) stellar-contamination-corrected transmission spectrum, as different stellar contamination parameters would be required to describe each visit for a joint fit. We perform free chemistry retrievals tailored to probe thin, potentially high mean molecular weight atmospheres of rocky planets which we had previously applied to TRAPPIST-1 c (M. Radica et al. 2025), and TRAPPIST-1 g (B. Benneke et al. 2025, submitted).

In SCARLET, the forward model iterates over the radiative transfer and hydrostatic equilibrium calculations for each sampled atmospheric composition and temperature structure to find the best-matching planetary radius at 10 mbar. Overall, our suite of retrievals included H_2 , He, N_2 , CH_4 , H_2O , CO, CO_2 , NH_3 , and SO_2 . For absorbing species, we use the HELIOS-K computed cross sections of H_2O (O. L. Polyansky et al. 2018), CO (R. J. Hargreaves et al. 2019), CO_2 (S. N. Yurchenko et al. 2020), CH_4 (R. J. Hargreaves et al. 2020), NH_3 (P. A. Coles et al. 2019), and SO_2 (D. S. Underwood et al. 2016). Our forward models are computed at $R = 31,250$, and binned to the resolution of the data for the likelihood calculation within our Bayesian inference framework.

We explored a range of parameterizations for the atmospheric composition to explain the spectrum of TRAPPIST-1 d. For one set of models, we consider single-molecule atmospheres (100% CH_4 , H_2O , CO_2 , CO, or NH_3) where we fit for the effective surface pressure P_{surf} (which could be the bare-rock surface or the top of an opaque cloud deck). We also test scenarios where N_2 is combined with one of CH_4 or CO_2 (fitting for the partial pressure of each component), as seen in the atmospheres of some solar system terrestrials. Finally, we also calculate models where all molecules (CH_4 , CO_2 , CO, NH_3 , H_2O , and SO_2) are together in a multigas atmosphere with H_2 as the background gas. We do not consider more complex models including aerosols with a variable size distribution, as we do not detect significant wavelength-dependent trends in our data. For the temperature profile, we fit for the temperature of the photosphere at the terminator, which is representative of the region probed by our transmission observations. We explore the parameter space using nested sampling, as implemented in the Python module *nestle* (J. Skilling 2004, 2006) and use at least 3000 live points per retrieval in the final results we present.

We further calculate a set of seven representative forward models using SCARLET for visual comparison to our observed spectrum: six corresponding to cloud-free 1 bar atmospheres fully composed of one of the absorbers we tested for (CH_4 , CO, H_2O , NH_3 , CO_2 , or SO_2) and one representative of a cloud-free, solar-composition atmosphere in chemical equilibrium (composition from J. W. Stock et al. 2018).

Table D1

Constraints on the Atmospheric Composition of TRAPPIST-1 d from the SCARLET Retrievals for Cases with High Atmospheric Mean Molecular Weight

| Atmospheric Scenario Composition | Sequential Fit $\log_{10}P_{\text{surf,eff}}$ (bar) | Joint Fit |
|-------------------------------------|--|-----------|
| 100% CH ₄ | <-4.66 | -4.77 |
| 100% CO | <-2.55 | -3.24 |
| 100% H ₂ O | <-2.42 | <-2.06 |
| 100% NH ₃ | <-3.09 | N/A |
| 100% CO ₂ | <0.52 | <-1.44 |
| 100% SO ₂ | <0.32 | <-1.95 |

Note. For each of the the single-molecule retrievals, we report the upper limit at 95% confidence obtained on the effective surface pressure in either the sequential fit or the joint stellar contamination + planetary atmosphere retrieval setup.

D.2. Joint Retrievals of Stellar Contamination and Planetary Atmosphere Contribution

The atmosphere-only retrievals presented above provide a first set of “sequential” retrievals, where the planetary atmosphere contribution is constrained from retrievals performed on the transmission spectrum of TRAPPIST-1 d after it has been corrected for the stellar contamination contribution constrained by *stctm*. We also perform more computationally expensive joint retrievals where we fit simultaneously for the contribution of unocculted stellar heterogeneities, and of a potential planetary atmosphere, to the observed spectrum of each visit. This second scheme is more conservative, as it marginalizes over the interplay between stellar and planetary contributions to the transmission spectrum instead of only using one realization of the stellar contamination signal. Further, planetary and stellar features could effectively “cancel out” to form a flat spectrum, which means that the data could allow for a broader range of atmosphere scenarios than afforded by the stellar-contamination-corrected spectrum.

In these joint retrievals, for each visit we fit the covering fractions of the cool and hot heterogeneities, and the differences between the photosphere temperature and that of each of the two heterogeneity populations. Informed by the TLS-only and out-of-transit spectral retrieval results, we fit the stellar $\log g$ but do not fit any $\log g$ difference between the photosphere and heterogeneity component. We also fit for an offset to account for a potential difference in the broadband transit depth between the two visits that could be introduced from leftover flux from the in-transit flare in visit 1 after the transit, affecting the baseline and acting as a dilution factor (as seen in M. Radica et al. 2025). This results in a total of 11 additional fitted parameters.

We use uniform and uninformative priors on each parameter, except for the photosphere temperature where we use the same Gaussian prior as in the *stctm* fit.

D.3. Impact of Refraction on the Sensitivity to Planetary Atmospheres

Atmospheric refraction occurs when starlight passes through an exoplanet’s atmosphere and bends due to gradients in the refractive index. In transmission spectroscopy, this bending of light can impose a refraction limit, effectively setting a boundary below which stellar light is refracted away from the

observer’s line of sight (A. Misra et al. 2014; T. D. Robinson et al. 2017; Y. Bétrémieux & M. R. Swain 2018). As a result, only the higher, less dense layers of the atmosphere contribute to the transmission spectrum. This phenomenon can restrict our ability to probe the deeper atmospheric regions, and is therefore crucial to consider when interpreting transmission spectra.

We use the Planetary Spectrum Generator (PSG; G. L. Villanueva et al. 2018, 2022) to compute the refraction limit from 3D GCM simulations of TRAPPIST-1 d’s atmospheric conditions (M. Turbet et al. 2023). PSG computes the refraction accounting for a nonuniform spherical atmosphere through numerical ray tracing. The light is traced from the observer through the planet’s atmosphere to the surface, or the top of the atmosphere, modulated by some adjustments of the incidence angles at each layer’s interface as a function of the refractive properties of the layers. The refraction coefficient is influenced by the layer composition, local density, and wavelength of the radiation. Since our spectrum rules out thick atmospheres for hydrogen-dominated atmospheres (Figure 3), we focus on potential secondary atmosphere compositions rich in volatiles, where a broader range of surface pressures can be invoked. We select two edge-case atmospheric scenarios in terms of the atmospheric mean molecular weight since it correlates with the refractive index (higher for heavier species): a water-dominated (10 bar H₂O, 1 bar N₂) composition, and a pure-CO₂ (10 bar CO₂) composition. The surface pressures in these models lie at the upper bound of our priors for the retrievals that explore high mean molecular weight atmospheres specifically (Figures 4 and 5). This test allows us to assess whether the limits we place across all the atmospheric scenarios we explore indeed lie within the range of pressures we would be sensitive to.

D.4. High Mean Molecular Weight Atmosphere Retrievals

We report the constraints obtained from single-molecule retrievals Table D1.

Appendix E

Water Loss Simulations for TRAPPIST-1 d, e, f, and g

We use the box model described in K. Moore et al. (2023), including the planetary mass dependence outlined in K. Moore et al. (2024), to assess the vulnerability of TRAPPIST-1 d to desiccation as a function of its initial water inventory. The model assumes that the planet remains at the same orbital distance throughout its evolution. The planet experiences time-variable water loss modulated by the amount of stellar irradiation it receives and by its evolving internal state, which dictates the amount of surface water susceptible to escape.

The stellar XUV irradiation is assumed to follow evolutionary models for low-mass stars (I. Baraffe et al. 2015). We model the XUV luminosity as initially saturated for 1 Gyr (I. Ribas et al. 2005), and account for a 5 Myr delay between star and planet formation, offsetting the stellar tracks accordingly to account for the cumulative XUV flux received by the planetary atmosphere.

The initial water inventory of the planet begins completely dissolved within the magma ocean. As the magma ocean solidifies, water is partitioned between three reservoirs: the magma ocean, the solidified mantle, and the planet surface (where, due to runaway greenhouse surface temperatures, water exists in a steam atmosphere). Due to the high solubility

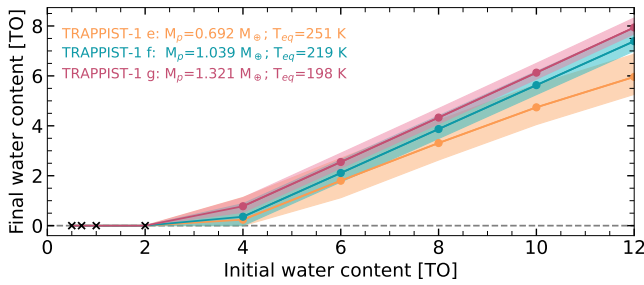


Figure E1. Water inventory evolution simulation outcomes for TRAPPIST-1 e, f, and g. The final global water inventories are shown for TRAPPIST-1 e (orange), f (blue), and g (pink), for the median stellar age of 7.6 Gyr (solid lines) and for the envelope of solutions over the range of stellar ages from 5.4 to 9.8 Gyr (semitransparent shaded regions). The horizontal dashed line represents the limit of desiccation, where a simulation for a planet that ends up completely devoid of water would lie (black crosses). For planets e, f, and g, in our simulations, water inventories of more than ~ 4 Earth oceans (“TO”) are enough for the planet to avoid complete desiccation to this day. The quoted planet mass equilibrium temperature values are taken from E. Agol et al. (2021).

of water in the initially fully molten mantle, a large fraction of the water is partitioned into the melt until the “bottom-up” solidification of the mantle is complete (see, e.g., L. T. Elkins-Tanton & S. Seager 2008; C. Dorn & T. Lichtenberg 2021). The complete solidification of the mantle coincides with the end of the runaway greenhouse phase, which is set by the level of irradiation received by the planet. Once the mantle is solidified, we model deep water cycling—where surface temperatures are now low enough that water can exist in condensed form on the surface—similar to that experienced on Earth due to plate tectonics (P. J. McGovern & G. Schubert 1989) over geological timescales. We note that the magma ocean could instead solidify before the end of the runaway gas accretion phase (see F. Selsis et al. 2023), but our simple model does not account for nonadiabatic thermal profiles that result in this early mantle solidification.

During the runaway greenhouse phase, water loss to space is modeled following energy-limited escape (A. J. Watson et al. 1981; N. V. Erkaev et al. 2007; E. D. Lopez 2017). The escape regime transitions to diffusion-limited escape (D. M. Hunten 1973) once surface water condensation becomes possible—with hydrogen slowly diffusing to the exobase where it can escape to space.

We model the evolution of a planet with the mass of TRAPPIST-1 d (Table A1) at several fixed orbital distances ranging from 0.2229 to 0.0526 au (to account for the uncertainty in the formation location and migration timescale of the planet) and with varying initial amounts of volatiles ranging from about 0.75 to 8 Earth oceans, which is reasonable for TRAPPIST-1 d (J. Krissansen-Totton & J. J. Fortney 2022). We model the water partitioning, water loss, and interior evolution of the planet over 5.4 to 9.8 Gyr, to account for the uncertainty on the stellar age (A. J. Burgasser & E. E. Mamajek 2017). Additionally, we investigate the water loss outcomes for TRAPPIST-1 e, f, and g, assuming initial volatile inventories ranging from 0.5 to 12 Earth oceans and assuming they remained at their present orbital distance for 5.4 to 9.8 Gyr of evolution.

The results for TRAPPIST-1 d are shown in Figure 7. We present the results of our simulations in terms of the final (surface + atmosphere) water inventories on TRAPPIST-1 e, f, and g assuming in situ evolution, for various initial volatile endowments (Figure E1).

ORCID iDs

- Caroline Piaulet-Ghorayeb <https://orcid.org/0000-0002-2875-917X>
 Björn Benneke <https://orcid.org/0000-0001-5578-1498>
 Martin Turbet <https://orcid.org/0000-0003-2260-9856>
 Pierre-Alexis Roy <https://orcid.org/0000-0001-6809-3520>
 Olivia Lim <https://orcid.org/0000-0003-4676-0622>
 René Doyon <https://orcid.org/0000-0001-5485-4675>
 Thomas J. Fauchez <https://orcid.org/0000-0002-5967-9631>
 Loïc Albert <https://orcid.org/0000-0003-0475-9375>
 Michael Radica <https://orcid.org/0000-0002-3328-1203>
 Louis-Philippe Coulombe <https://orcid.org/0000-0002-2195-735X>
 David Lafrenière <https://orcid.org/0000-0002-6780-4252>
 Nicolas B. Cowan <https://orcid.org/0000-0001-6129-5699>
 Danika Belzile <https://orcid.org/0009-0007-6298-9802>
 Kamrul Musfirat <https://orcid.org/0009-0007-6653-6165>
 Mehramat Kaur <https://orcid.org/0009-0004-9827-2481>
 Alexandrine L’Heureux <https://orcid.org/0009-0005-6135-6769>
 Doug Johnstone <https://orcid.org/0000-0002-6773-459X>
 Ryan J. MacDonald <https://orcid.org/0000-0003-4816-3469>
 Romain Allart <https://orcid.org/0000-0002-1199-9759>
 Lisa Dang <https://orcid.org/0000-0003-4987-6591>
 Lisa Kaltenegger <https://orcid.org/0000-0002-0436-1802>
 Stefan Pelletier <https://orcid.org/0000-0002-8573-805X>
 Jason F. Rowe <https://orcid.org/0000-0002-5904-1865>
 Jake Taylor <https://orcid.org/0000-0003-4844-9838>
 Jake D. Turner <https://orcid.org/0000-0001-7836-1787>

References

- Acuña, L., Deleuil, M., Mousis, O., et al. 2021, *A&A*, 647, A53
 Agol, E., Dorn, C., Grimm, S. L., et al. 2021, *PSJ*, 2, 1
 Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, 558, A33
 Baraffe, I., Homeier, D., Allard, F., & Chabrier, G. 2015, *A&A*, 577, A42
 Barr, A. C., Dobos, V., & Kiss, L. L. 2018, *A&A*, 613, A37
 Beeck, B., Schüssler, M., Cameron, R. H., & Reiners, A. 2015, *A&A*, 581, A42
 Bell, T., Ahrer, E.-M., Brande, J., et al. 2022, *JOSS*, 7, 4503
 Benneke, B. 2013, AAS Meeting, 221, 224.04
 Benneke, B. 2015, arXiv, arXiv:1504.07655
 Benneke, B., Knutson, H. A., Lothringer, J., et al. 2019a, *NatAs*, 3, 813
 Benneke, B., & Seager, S. 2012, AAS Meeting, 44, 103.06
 Benneke, B., Werner, M., Petigura, E., et al. 2017, *ApJ*, 834, 187
 Benneke, B., Wong, I., Piaulet, C., et al. 2019b, *ApJL*, 887, L14
 Berardo, D., de Wit, J., Gillon, M., et al. 2025, arXiv, arXiv:2506.12140
 Berta, Z. K., Charbonneau, D., Bean, J., et al. 2011, *ApJ*, 736, 12
 Bétrémieux, Y., & Swain, M. R. 2018, *ApJ*, 865, 12
 Birkmann, S. M., Ferruit, P., Giardini, G., et al. 2022, *A&A*, 661, A83
 Bolmont, E., Selsis, F., Owen, J. E., et al. 2017, *MNRAS*, 464, 3728
 Bower, D. J., Hakim, K., Sossi, P. A., & Sanan, P. 2022, Retention of Water in Terrestrial Magma Oceans and Carbon-rich Early Atmospheres, *PSJ*, 3, 93
 Bullock, M. A., & Grinspoon, D. H. 2013, in Comparative Climatology of Terrestrial Planets, ed. S. J. Mackwell et al. (Tucson: Univ. of Arizona Press), 19
 Burgasser, A. J., & Mamajek, E. E. 2017, *ApJ*, 845, 110
 Cadieux, C., Doyon, R., MacDonald, R. J., et al. 2024, *ApJ*, 970, L2
 Catling, D. C., & Zahnle, K. J. 2020, *SciA*, 6, eaax1420
 Charnay, B., Forget, F., Tobie, G., Sotin, C., & Wordsworth, R. 2014, *Icar*, 241, 269
 Coleman, G. A. L., Leleu, A., Alibert, Y., & Benz, W. 2019, *A&A*, 631, A7
 Coles, P. A., Yurchenko, S. N., & Tennyson, J. 2019, *MNRAS*, 490, 4638
 Coulombe, L.-P., Radica, M., Benneke, B., et al. 2025, *NatAS*, 9, 512
 Coulombe, L.-P., Roy, P.-A., & Benneke, B. 2024, *AJ*, 168, 227
 Cyr, K. E., Sears, W. D., & Lunine, J. I. 1998, *Icar*, 135, 537
 Davoudi, F., Rackham, B. V., Gillon, M., et al. 2024, *ApJL*, 970, L4

- de Wit, J., Wakeford, H. R., Lewis, N. K., et al. 2018, *NatAs*, **2**, 214
- Dorn, C., & Lichtenberg, T. 2021, *ApJL*, **922**, L4
- Doshi, D., Cowan, N. B., & Huang, Y. 2022, *MNRAS*, **515**, 1982
- Doyon, R. 2024, arXiv:2403.12617
- Ducrot, E., Gillon, M., Delrez, L., et al. 2020, *A&A*, **640**, A112
- Ducrot, E., Lagage, P.-O., Min, M., et al. 2024, *NatAs*, **9**, 358
- Ducrot, E., Sestovic, M., Morris, B. M., et al. 2018, *AJ*, **156**, 218
- Elkins-Tanton, L. T., & Seager, S. 2008, *ApJ*, **685**, 1237
- Encrenaz, T., & Coustenis, A. 2018, in *Handbook of Exoplanets*, ed. H. J. Deeg & J. A. Belmonte (Berlin: Springer)
- Erkaev, N. V., Kulikov, Y. N., Lammer, H., et al. 2007, *A&A*, **472**, 329
- Faucher, T. J., Turbet, M., Villanueva, G. L., et al. 2019, *ApJ*, **887**, 194
- Foreman-Mackey, D. 2016, *JOSS*, **1**, 24
- Foreman-Mackey, D., Agol, E., Ambikasaran, S., & Angus, R. 2017, *AJ*, **154**, 220
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, **125**, 306
- Fournier-Tondreau, M., MacDonald, R. J., Radica, M., et al. 2023, *MNRAS*, **528**, 3354
- Garcia, L. J., Moran, S. E., Rackham, B. V., et al. 2022, *A&A*, **665**, A19
- Gialluca, M. T., Barnes, R., Meadows, V. S., et al. 2024, *PSJ*, **5**, 137
- Gillon, M., Triaud, A. H. M. J., Demory, B.-O., et al. 2017, *Natur*, **542**, 456
- Grant, D., & Wakeford, H. R. 2024, *JOSS*, **9**, 6816
- Grayver, A., Bower, D. J., Saur, J., Dorn, C., & Morris, B. M. 2022, *ApJL*, **941**, L7
- Greene, T. P., Bell, T. J., Ducrot, E., et al. 2023, *Natur*, **618**, 39
- Grimm, S. L., Demory, B.-O., Gillon, M., et al. 2018, *A&A*, **613**, A68
- Hargreaves, R. J., Gordon, I. E., Rey, M., et al. 2020, *ApJS*, **247**, 55
- Hargreaves, R. J., Gordon, I. E., Rothman, L. S., et al. 2019, *JQSRT*, **232**, 35
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Natur*, **585**, 357
- Herbert, O., Woitke, P., Helling, C., & Zerkle, A. L. 2022, *A&A*, **658**, A180
- Hill, M. L., Bott, K., Dalba, P. A., et al. 2023, *AJ*, **165**, 34
- Hori, Y., & Ogihara, M. 2020, *ApJ*, **889**, 77
- Howard, W. S., Kowalski, A. F., Flagg, L., et al. 2023, *ApJ*, **959**, 64
- Hu, R., Gaillard, F., & Kite, E. S. 2023, *ApJL*, **948**, L20
- Hu, R., & Thomas, T. B. 2022, *NatGe*, **15**, 106
- Hunten, D. M. 1973, *JatS*, **30**, 1481
- Hunter, J. D. 2007, *CSE*, **9**, 90
- Husser, T.-O., Wende-von Berg, S., Dreizler, S., et al. 2013, *A&A*, **553**, A6
- Ih, J., Kempton, E. M. R., Whittaker, E. A., & Lessard, M. 2023, *ApJL*, **952**, L4
- Ikoma, M., Elkins-Tanton, L., Hamano, K., & Suckale, J. 2018, *SSRv*, **214**, 76
- Iyer, A. R., & Line, M. R. 2020, *ApJ*, **889**, 78
- Iyer, A. R., Line, M. R., Muirhead, P. S., Fortney, J. J., & Gharib-Nezhad, E. 2023a, *ApJ*, **944**, 41
- Iyer, A. R., Line, M. R., Muirhead, P. S., Fortney, J. J., & Gharib-Nezhad, E. 2023b, The SPHINX M-dwarf Spectral Grid. I. Benchmarking New Model Atmospheres to Derive Fundamental M-dwarf Properties, v4.0.0., doi:10.5281/zenodo.11392341
- Jahandar, F., Doyon, R., Artigau, et al. 2025, *ApJ*, **978**, 154
- Kasai, Y., Dupuy, E., Saito, R., et al. 2011, *ACP*, **11**, 8607
- Kelsey, V., Riley, S., & Minschwaner, K. 2022, *AMT*, **15**, 1563
- Kipping, D. M. 2013, *MNRAS*, **435**, 2152
- Kite, E. S. 2019, *SSRv*, **215**, 10
- Komacek, T. D., Faucher, T. J., Wolf, E. T., & Abbot, D. S. 2020, *ApJL*, **888**, L20
- Kopparapu, R., Wolf, E. T., Arney, G., et al. 2017, *ApJ*, **845**, 5
- Kopparapu, R. K., Ramirez, R., Kasting, J. F., et al. 2013, *ApJ*, **765**, 131
- Kopparapu, R. K., Wolf, E. T., Haqq-Misra, J., et al. 2016, *ApJ*, **819**, 84
- Kowalski, A. F., Hawley, S. L., Wisniewski, J. P., et al. 2013, *ApJS*, **207**, 15
- Kreidberg, L. 2015, *PASP*, **127**, 1161
- Krissansen-Totton, J. 2023, *ApJL*, **951**, L39
- Krissansen-Totton, J., & Fortney, J. J. 2022, *ApJ*, **933**, 115
- Krissansen-Totton, J., Olson, S., & Catling, D. C. 2018, *SciA*, **4**, eaao5747
- Lichtenberg, T., & Miguel, Y. 2025, *TrGeo*, **7**, 51
- Liebert, J., & Gizis, J. E. 2006, *PASP*, **118**, 659
- Liggins, P., Jordan, S., Rimmer, P. B., & Shorttle, O. 2022, *JGRE*, **127**, e07123
- Lim, O., Benneke, B., Doyon, R., et al. 2023, *ApJ*, **955**, L22
- Lincowski, A. P., Meadows, V. S., Crisp, D., et al. 2018, *ApJ*, **867**, 76
- Lincowski, A. P., Meadows, V. S., Zieba, S., et al. 2023, *ApJ*, **955**, 7
- Lopez, E. D. 2017, *MNRAS*, **472**, 245
- Luger, R., Sestovic, M., Kruse, E., et al. 2017, *NatAs*, **1**, 0129
- Lustig-Yaeger, J., Meadows, V. S., & Lincowski, A. P. 2019, *AJ*, **158**, 27
- Mann, A. W., Dupuy, T., Kraus, A. L., et al. 2019, *ApJ*, **871**, 63
- Massol, H., Davaille, A., & Sarda, P. 2023, *JGRE*, **128**, e2023JE007848
- McGovern, P. J., & Schubert, G. 1989, *E&PSL*, **96**, 27
- Meadows, V. S., Lincowski, A. P., & Lustig-Yaeger, J. 2023, *PSJ*, **4**, 192
- Meadows, V. S., Reinhard, C. T., Arney, G. N., et al. 2017, *AsBio*, **18**, 630
- Misra, A., Meadows, V., & Crisp, D. 2014, *ApJ*, **792**, 61
- Moore, K., Cowan, N. B., & Boukaré, C. 2023, *MNRAS*, **526**, 6235
- Moore, K., David, B., Zhang, A. Y., & Cowan, N. B. 2024, *ApJ*, **972**, 131
- Moran, S. E., Hörst, S. M., Batalha, N. E., Lewis, N. K., & Wakeford, H. R. 2018, *AJ*, **156**, 252
- Moran, S. E., Stevenson, K. B., Sing, D. K., et al. 2023, *ApJL*, **948**, L11
- Morley, C. V., Fortney, J. J., Marley, M. S., et al. 2015, *ApJ*, **815**, 110
- Morley, C. V., Knutson, H., Line, M., et al. 2017, *AJ*, **153**, 86
- Norris, C. M., Unruh, Y. C., Witzke, V., et al. 2023, *MNRAS*, **524**, 1139
- Pelletier, S., Benneke, B., Darveau-Bernier, A., et al. 2021, *AJ*, **162**, 73
- Piaulet, C., Benneke, B., Almenara, J. M., et al. 2023, *NatAs*, **7**, 206
- Piaulet-Ghorayeb, C. 2024, First Release of Stellar Contamination Modeling and Retrieval Code, v1.0.0, doi:10.5281/zenodo.13153252
- Piaulet-Ghorayeb, C., Benneke, B., Radica, M., et al. 2024, *ApJL*, **974**, L10
- Polyansky, O. L., Kyuberis, A. A., Zobov, N. F., et al. 2018, *JOSS*, **480**, 2597
- Rackham, B. V., Apai, D., & Giampapa, M. S. 2018, *ApJ*, **853**, 122
- Rackham, B. V., Espinoza, N., Berdyugina, S. V., et al. 2023, *RASTI*, **2**, 148
- Radica, M., Coulombe, L.-P., Taylor, J., et al. 2024, *ApJ*, **962**, L20
- Radica, M., Piaulet-Ghorayeb, C., Taylor, J., et al. 2025, *ApJL*, **979**, L5
- Rathcke, A. D., Buchhave, L. A., de Wit, J., et al. 2025, *ApJL*, **979**, L19
- Redfield, S., Batalha, N., Benneke, B., et al. 2024, arXiv:2404.02932
- Ribas, I., Guinan, E. F., Güdel, M., & Audard, M. 2005, *ApJ*, **622**, 680
- Robinson, T. D., Fortney, J. J., & Hubbard, W. B. 2017, *ApJ*, **850**, 128
- Rustamkulov, Z., Sing, D., Mukherjee, S., et al. 2023, AAS Meeting, **55**, 124.01
- Schmandt, B., Jacobsen, S. D., Becker, T. W., Liu, Z., & Dueker, K. G. 2014, *Sci*, **344**, 1265
- Seager, S., & Shapiro, A. I. 2024, *ApJ*, **970**, 155
- Selsis, F., Leconte, J., Turbet, M., Chaverot, G., & Bolmont, É. 2023, *Natur*, **620**, 287
- Sing, D. K., Pont, F., Aigrain, S., et al. 2011, *MNRAS*, **416**, 1443
- Skilling, J. 2004, *AIPC*, **735**, 395
- Skilling, J. 2006, *BayAn*, **1**, 833
- Smith, H. N., Shapiro, A. I., Witzke, V., et al. 2025, *ApJL*, **978**, L13
- Stock, J. W., Kitzmann, D., Patzer, A. B. C., & Sedlmayr, E. 2018, *MNRAS*, **479**, 865
- Townsend, R., & Lopez, A. 2023, *JOSS*, **8**, 4602
- TRAPPIST-1 JWST Community Initiative, de Wit, J., Doyon, R., et al. 2024, *NatAs*, **8**, 810
- Turbet, M., Bolmont, E., Bourrier, V., et al. 2020, *SSRv*, **216**, 100
- Turbet, M., Bolmont, E., Leconte, J., et al. 2018, *A&A*, **612**, A86
- Turbet, M., & Forget, F. 2021, arXiv, arXiv:2103.10301
- Turbet, M., Faucher, T. J., Leconte, J., et al. 2023, *A&A*, **679**, A126
- Underwood, D. S., Tennyson, J., Yurchenko, S. N., et al. 2016, *MNRAS*, **459**, 3890
- Unterborn, C. T., Desch, S. J., Hinkel, N. R., & Lorenzo, A. 2018, *NatAs*, **2**, 297
- Van Looveren, G., Güdel, M., Boro Saikia, S., & Kislyakova, K. 2024, *A&A*, **683**, A153
- Villanueva, G. L., Liuzzi, G., Faggi, S., et al. 2022, Fundamentals of the Planetary Spectrum Generator (Washington, DC: NASA)
- Villanueva, G. L., Smith, M. D., Protopapa, S., Faggi, S., & Mandell, A. M. 2018, *JQSRT*, **217**, 86
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *NatMe*, **17**, 261
- Wakeford, H. R., Lewis, N. K., Fowler, J., et al. 2019, *AJ*, **157**, 11
- Wang, S., Wu, D.-H., Barclay, T., & Laughlin, G. P. 2017, arXiv, arXiv:1704.04290
- Watson, A. J., Donahue, T. M., & Walker, J. C. G. 1981, *Icar*, **48**, 150
- Way, M. J. 2025, *ApJ*, **980**, L7
- Wolf, E. T. 2017, *ApJ*, **839**, L1
- Yang, J., Cowan, N. B., & Abbot, D. S. 2013, *ApJL*, **771**, L45
- Yurchenko, S. N., Mellor, T. M., Freedman, R. S., & Tennyson, J. 2020, *MNRAS*, **496**, 5282
- Zahnle, K. J., & Catling, D. C. 2017, *ApJ*, **843**, 122
- Zhang, Z., Zhou, Y., Rackham, B. V., & Apai, D. 2018, *AJ*, **156**, 178
- Zieba, S., Kreidberg, L., Ducrot, E., et al. 2023, *Natur*, **620**, 746